GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:22:15 ; Search time 50.86 Seconds
(without alignments)
3.735 Million cell updates/sec

Title: US-09-744-804-35
Perfect score: 53
Sequence: 1 KQGNFNAWV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep *
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep *
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep *
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep *
5: /cgn2_6/ptodata/2/iaa/backfiles1.pep *
6: /cgn2_6/ptodata/2/iaa/backfiles2.pep *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	53	100.0	159	2	US-08-162-402B-12	Sequence 12, Appl
2	53	100.0	217	1	US-07-607-538C-3	Sequence 3, Appli
3	53	100.0	217	2	US-08-162-402B-3	Sequence 3, Appli
4	53	100.0	218	1	US-07-607-538C-2	Sequence 2, Appli
5	53	100.0	218	2	US-08-162-402B-2	Sequence 2, Appli
6	53	100.0	387	2	US-08-162-402B-6	Sequence 6, Appli
7	53	100.0	465	2	US-08-162-402B-8	Sequence 8, Appli
8	39	73.6	369	2	US-08-424-224-2	Sequence 2, Appli
9	39	73.6	369	5	PCT-US94-02891-69	Sequence 69, Appl
10	36	67.9	321	2	US-08-480-229C-21	Sequence 21, Appl
11	36	67.9	321	2	US-08-659-235C-21	Sequence 21, Appl
12	36	67.9	480	2	US-08-480-229C-10	Sequence 10, Appl
13	36	67.9	480	2	US-08-659-235C-10	Sequence 10, Appl
14	36	67.9	513	2	US-08-480-229C-14	Sequence 14, Appl
15	36	67.9	513	2	US-08-659-235C-14	Sequence 14, Appl
16	34	64.2	347	1	US-08-052-205-7	Sequence 7, Appli
17	34	64.2	347	1	US-08-595-974-7	Sequence 7, Appli
18	34	64.2	369	1	US-08-052-205-4	Sequence 4, Appli
19	34	64.2	369	1	US-08-595-974-4	Sequence 4, Appli
20	33	62.3	227	1	US-08-458-023B-4	Sequence 4, Appli
21	33	62.3	229	2	US-08-121-436A-4	Sequence 4, Appli
22	33	62.3	1041	3	US-08-008-978-2	Sequence 2, Appli
23	32	60.4	157	2	US-08-162-402B-13	Sequence 13, Appl
24	32	60.4	320	2	US-08-480-229C-20	Sequence 20, Appl
25	32	60.4	320	2	US-08-659-235C-20	Sequence 20, Appl
26	32	60.4	404	2	US-08-082-197C-62	Sequence 62, Appl
27	32	60.4	464	2	US-08-162-402B-9	Sequence 9, Appli
28	32	60.4	477	1	US-08-191-338A-2	Sequence 2, Appli
29	32	60.4	1121	1	US-07-799-915A-2	Sequence 2, Appli
30	32	60.4	1121	1	US-08-335-022C-2	Sequence 2, Appli
31	32	60.4	1121	1	US-08-487-203A-2	Sequence 2, Appli
32	32	60.4	1334	6	5476657-1	Patent No. 5476657
33	32	60.4	1403	1	US-07-908-253-3	Sequence 3, Appli
34	32	60.4	1403	2	US-08-694-865-17	Sequence 17, Appl
35	32	60.4	1403	2	US-08-535-837-3	Sequence 3, Appli
36	32	60.4	1403	3	US-09-124-491-17	Sequence 17, Appl
37	32	60.4	1541	4	US-08-296-791-3	Sequence 3, Appli
38	32	60.4	1541	5	PCT-US95-10661A-3	Sequence 3, Appli
39	32	60.4	1545	4	US-08-296-791-4	Sequence 4, Appli
40	32	60.4	1545	5	PCT-US95-10661A-4	Sequence 4, Appli
41	32	60.4	1702	4	US-08-296-791-5	Sequence 5, Appli
42	32	60.4	1702	5	PCT-US95-10661A-5	Sequence 5, Appli
43	31	58.5	107	1	US-08-036-210-4	Sequence 4, Appli
44	31	58.5	107	2	US-08-449-609-4	Sequence 4, Appli
45	31	58.5	159	2	US-08-162-402B-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-162-402B-12
; Sequence 12, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAPORCA, DAVID J.
; TITLE OF INVENTION: 4E EPALRON HUMAN MILK FAT
; TITLE OF INVENTION: GLYCOLIPID (HMFG) ANTIGEN, PEANUTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amsel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P&S 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-489-1600 7700
; TELEFAX: 213-489-4210
; TELEX
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-12

Query Match 100.0%; Score 53; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 0.034;

```
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
           |||||||||
Db      43 KQGNFNAWV 51


RESULT   2
US-07-607-538C-3
; Sequence 3, Application US/07607538C
; Patent No. 5455031
;  GENERAL INFORMATION:
;    APPLICANT:  Ceriani Dr., Roberto L.
;    APPLICANT:  Peterson Dr., Jerry A.
;    APPLICANT:  Larocca, David J.
;    TITLE OF INVENTION:  POLYPEPTIDE WITH 46
;    TITLE OF INVENTION:  DIFFERENTIATION ANTIGEN BINDING SPECIFITY AND  CLOTTING
;    TITLE OF INVENTION:  FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
;    TITLE OF INVENTION:  FUSION   PROTEIN,  POLYNUCLEOTIDE AND  POLYRIBO-
;    TITLE OF INVENTION:  NUCLEOTIDE  ENCODING  THE POLYPEPTIDE, ANTI-
;    TITLE OF INVENTION:  POLYPEPTIDE ANTIBODIES, KITS  AND  METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  V.  Amzel & Assoc.
;      STREET:  2055 No. 5455031th Broadway
;      CITY:  Walnut Creek
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94596
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS 5.0
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/607,538C
;      FILING DATE:  01 NOV 1990
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Viviana Amzel
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  CRFCC-004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 943-1931
;      TELEFAX:  (510) 943-1189
;      TELEX:  N.A.
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  217 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:
US-07-607-538C-3


  Query Match             100.0%,  Score 53,  DB 1,  Length 217;
  Best Local Similarity   100.0%;  Pred. No. 0.047;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
           |||||||||
Db     101 KQGNFNAWV 109


RESULT   3
US-08-162-402B-3
; Sequence 4, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  217 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-3


  Query Match             100.0%,  Score 53,  DB 2,  Length 217;
  Best Local Similarity   100.0%;  Pred. No. 0.047;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
           |||||||||
Db     101 KQGNFNAWV 109


RESULT   4
US-07-607-538C-2
; Sequence 2, Application US/07607538C
; Patent No. 5455031
;  GENERAL INFORMATION:
;    APPLICANT:  Ceriani Dr., Roberto L.
;    APPLICANT:  Peterson Dr., Jerry A.
;    APPLICANT:  Larocca, David J.
;    TITLE OF INVENTION:  POLYPEPTIDE WITH 46
;    TITLE OF INVENTION:  DIFFERENTIATION ANTIGEN BINDING SPECIFITY AND  CLOTTING
;    TITLE OF INVENTION:  FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
;    TITLE OF INVENTION:  FUSION   PROTEIN,  POLYNUCLEOTIDE AND  POLYRIBO
;    TITLE OF INVENTION:  NUCLEOTIDE  ENCODING  THE POLYPEPTIDE, ANTI-
;    TITLE OF INVENTION:  POLYPEPTIDE ANTIBODIES, KITS  AND  METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  V.  Amzel & Assoc.
;      STREET:  2055 No. 5455031th Broadway
;      CITY:  Walnut Creek
;      STATE:  California
;      COUNTRY:  USA
```

```
;       ZIP:  94596
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS 5.0
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/607,538C
;       FILING DATE:  01-NOV-1990
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Viviana Amzel
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  CRFCC-004
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (510) 943-1931
;       TELEFAX:  (510) 943-1189
;       TELEX:  N A
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  218 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     FRAGMENT TYPE:
US-07-607-538C-2


  Query Match             100.0%;  Score 53;  DB 1;  Length 218;
  Best Local Similarity   100.0%;  Pred. No. 0.047;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
          |||||||||
Db     102 KQGNFNAWV 110


RESULT   5
US-08-162-402B-2
; Sequence 2, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;     APPLICANT:  CERIANI, ROBERTO L.
;     APPLICANT:  PETERSON, JERRY A.
;     APPLICANT:  LAROCCA, DAVID J.
;     TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;     TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;     NUMBER OF SEQUENCES:  29
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pretty, Schroeder & Poplawski
;       STREET:  444 South Flower St., 19th Floor
;       CITY:  Los Angeles
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/162,402B
;       FILING DATE:  03-DEC-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Amzel, Viviana
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  P&G 28215
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  213-622-7700
;       TELEFAX:  213-489-4210
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  218 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
US-08-162-402B-2


  Query Match             100.0%;  Score 53;  DB 2;  Length 218;
  Best Local Similarity   100.0%;  Pred. No. 0.047;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
          |||||||||
Db     102 KQGNFNAWV 110


RESULT   6
US-08-162-402B-6
; Sequence 6, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;     APPLICANT:  CERIANI, ROBERTO L.
;     APPLICANT:  PETERSON, JERRY A.
;     APPLICANT:  LAROCCA, DAVID J.
;     TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;     TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;     NUMBER OF SEQUENCES:  29
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pretty, Schroeder & Poplawski
;       STREET:  444 South Flower St., 19th Floor
;       CITY:  Los Angeles
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/162,402B
;       FILING DATE:  03-DEC-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Amzel, Viviana
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  P&G 28215
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  213-622-7700
;       TELEFAX:  213-489-4210
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  387 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
US-08-162-402B-6


  Query Match             100.0%;  Score 53;  DB 2;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.084;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 KQGNFNAWV 9
           |||||||||
Db     271 KQGNFNAWV 279


RESULT   7
US 08-162-402B-8
; Sequence 8, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;     APPLICANT:  CERIANI, ROBERTO L.
;     APPLICANT:  PETERSON, JERRY A.
;     APPLICANT:  LAROCCA, DAVID J.
;     TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;     TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;     NUMBER OF SEQUENCES:  29
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pretty, Schroeder & Poplawski
;       STREET:  444 South Flower St., 19th Floor
;       CITY:  Los Angeles
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/162,402B
;       FILING DATE:  03-DEC-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Amzel, Viviana
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  P66 38215
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  213-622-7700
;       TELEFAX:  213-489-4210
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  465 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
US-08-162-402B-8


  Query Match             100.0%;  Score 53;  DB 2;  Length 465;
  Best Local Similarity   100.0%;  Pred. No. 0.1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
           |||||||||
Db     349 KQGNFNAWV 357


RESULT   8
US-08-424-224-2
; Sequence 2, Application US/08424224
; Patent No. 5912173
;  GENERAL INFORMATION:
;     APPLICANT:  LEONARD, WARREN J.
;     TITLE OF INVENTION:  MURINE IL-2R  cDNA AND
;     TITLE OF INVENTION:  USES THEREOF
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MORGAN & FINNEGAN
;       STREET:  345 PARK AVE.
;       CITY:  NEW YORK
;       STATE:  NEW YORK
;       COUNTRY:  USA
;       ZIP:  10154
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  FLOPPY DISK
;       COMPUTER:  IBM PC COMPATIBLE
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WORD PERFECT # 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/424,224
;       FILING DATE:
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/121,435
;       FILING DATE:  14-SEPT-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  WILLIAM S. FEILER
;       REGISTRATION NUMBER:  26,728
;       REFERENCE/DOCKET NUMBER:  2026-4061US1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-758-4800
;       TELEFAX:  212-751-6849
;       TELEX:  421792
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  369
;       TYPE:  AMINO ACID
;       TOPOLOGY:  UNKNOWN
;     MOLECULE TYPE:
;       DESCRIPTION:  PROTEIN
;     HYPOTHETICAL:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  MURINE
;       INDIVIDUAL ISOLATE:  IL-2R
US 08 424 224-2


  Query Match             73.6%;  Score 39;  DB 2;  Length 369;
  Best Local Similarity   85.7%;  Pred. No. 20;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 QGNFNAW 8
           ||||:||
Db     304 QGNFSAW 310


RESULT   9
PCT-US94-02891-69
; Sequence 69, Application PC/TUS9402891
;  GENERAL INFORMATION:
;     APPLICANT:  THE GOVERNMENT OF THE UNITED STATES OF AMERICA AS
;     APPLICANT:  REPRESENTED BY THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN
;     APPLICANT:  SERVICES
;     APPLICANT:  OFFICE OF TECHNOLOGY TRANSFER, NATIONAL
;     APPLICANT:  INSTITUTES OF HEALTH, BOX OTT, BETHESDA, MARYLAND 20892  USA
;     TITLE OF INVENTION:  METHODS FOR DIAGNOSIS AND TREATMENT OF
;     TITLE OF INVENTION:  XSCID
;     NUMBER OF SEQUENCES:  69
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MORGAN & FINNEGAN
;       STREET:  345 PARK AVE.
;       CITY:  NEW YORK
;       STATE:  NEW YORK
;       COUNTRY:  USA
;       ZIP:  10154
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  FLOPPY DISK
;       COMPUTER:  IBM PC COMPATIBLE
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WORD PERFECT # 5.1
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/02891
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/031,143
;      FILING DATE:  12-MAR-1993
;      APPLICATION NUMBER:  08/121,435
;      FILING DATE:  14-SEPT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WILLIAM S  FEIER
;      REGISTRATION NUMBER:  26,728
;      REFERENCE/DOCKET NUMBER:  2026-4061
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-758-4800
;      TELEFAX:  212-751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  69:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  369
;      TYPE:  AMINO ACID
;      TOPOLOGY:  UNKNOWN
;    MOLECULE TYPE:
;      DESCRIPTION:  PROTEIN
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  MURINE
;      INDIVIDUAL ISOLATE:  IL-2R
PCT-US94-02891-69


  Query Match             73.6%;  Score 39;  DB 5;  Length 369;
  Best Local Similarity   85.7%;  Pred. No. 20;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 QGNFNAW 8
           ||||:||
Db     304 QGNFSAW 310


RESULT  10
US-08-480-229C-21
; Sequence 21, Application US/08480229C
; Patent No. 5874562
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H  Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS-1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036 2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,229C
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8907-0026-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  321 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-480-229C-21


  Query Match             67.9%;  Score 36;  DB 2;  Length 321;
  Best Local Similarity   75.0%;  Pred. No. 56;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAW 8
             |||  |||
Db     204 KQGKVNAW 211


RESULT  11
US-08-659-235C-21
; Sequence 21, Application US/08659235C
; Patent No. 5877281
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS-1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,235C
;      FILING DATE:  05-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8907-0026-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  321 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-659-235C-21


  Query Match             67.9%;  Score 36;  DB 2;  Length 321;
  Best Local Similarity   75.0%;  Pred. No. 56;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAW 8
```

```
          ||| |||
Db    204 KQGKVNAW 211


RESULT 12
US-08-480-229C-10
; Sequence 10, Application US/08480229C
; Patent No. 5874562
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS-1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,229C
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8907-0026-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  480 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-229C-10


  Query Match             67.9%;  Score 36;  DB 2;  Length 480;
  Best Local Similarity   75.0%;  Pred. No. 84;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAW 8
          ||| |||
Db     360 KQGKVNAW 367


RESULT 13
US-08-659-235C-10
; Sequence 10, Application US/08659235C
; Patent No. 5877281
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS-1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,235C
;      FILING DATE:  05-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8907-0034-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  480 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-659-235C-10


  Query Match             67.9%;  Score 36;  DB 2;  Length 480;
  Best Local Similarity   75.0%;  Pred. No. 84;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAW 8
          ||| |||
Db     360 KQGKVNAW 367


RESULT 14
US-08-480-229C-14
; Sequence 14, Application US/08480229C
; Patent No. 5874562
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS-1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,229C
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8907-0026-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
```

```
;       TELEFAX:  (212) 869 8864/9741
;       TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  513 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-229C-14


  Query Match             67.9%;  Score 36;  DB 2;  Length 513;
  Best Local Similarity   75.0%;  Pred. No. 89;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAW 8
           ||| ||| |
Db     393 KQGKVNAW 400


RESULT  15
US-08-659-235C-14
; Sequence 14, Application US/08659235C
; Patent No. 5877281
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Briqid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS-1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,235C
;      FILING DATE:  05-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8907-0034-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  513 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-659-235C-14


  Query Match             67.9%;  Score 36;  DB 2;  Length 513;
  Best Local Similarity   75.0%;  Pred. No. 89;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAW 8
           ||| ||| |
Db     393 KQGKVNAW 400


Search completed:  September  5, 2002, 15:27:33
Job time:  318 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

```
Run on:          September 5, 2002, 15:23:20 ; Search time 158.43 Seconds
                                            (without alignments)
                                            6.310 Million cell updates/sec

Title:           US-09-744-804-35
Perfect score:   53
Sequence:        1 KQCNFNAWV 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters        747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_032802:*
                 1:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
                 2:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
                 3:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
                 4:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
                 5:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
                 6:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
                 7:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
                 8:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
                 9:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
                 10:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
                 11:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
                 12:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
                 13:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
                 14:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
                 15:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
                 16:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
                 17:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
                 18:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
                 19:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
                 20:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
                 21:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
                 22:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	53	100.0	9	21	AAY82840	Lactadherin (BA-46
2	53	100.0	217	16	AAR77253	HMFG 46 kDa antige
3	53	100.0	218	16	AAR77254	HMFG 46 kDa antige
4	53	100.0	387	16	AAR77252	HMFG 46 kDa antige
5	53	100.0	387	21	AAY94453	Human lactadherin
6	39	73.6	363	22	AAB71682	Bos taurus interle
7	39	73.6	369	15	AAP59094	Murine IL-2R gamma
8	38	71.7	674	22	ABG29559	Novel human diagno
9	38	71.7	1087	22	ABG24996	Novel human diagno
10	36	67.9	51	22	AAM65565	Human brain expres
11	36	67.9	68	22	AAB65971	Human secreted pro
12	36	67.9	113	22	AAB94968	Rat zcytor 10 cyto
13	36	67.9	181	22	AAB71691	CRCGCL related pep
14	36	67.9	321	20	AAW91697	Human milk fat glo
15	36	67.9	357	22	AAB04966	Zcytor 10 cytokine
16	36	67.9	359	21	AAB23543	Murine cytokine re
17	36	67.9	359	22	AAB04973	Zcytor 10 cytokine
18	36	67.9	371	20	AAY31810	Human cytokine rec
19	36	67.9	371	22	AAB71681	CRCGCL protein. H
20	36	67.9	480	18	AAW10364	Mouse developmenta
21	36	67.9	480	20	AAW94683	Murine Del 1 prote
22	36	67.9	481	18	AAW10365	Human developmenta
23	36	67.9	513	20	AAW94685	Human Del-1 protei
24	36	67.9	1384	18	AAW31946	Human CASPR/p190,
25	35	66.0	80	22	AAG02699	Human polypeptide
26	35	66.0	141	22	ABG05534	Novel human diagno
27	35	66.0	228	21	AAG19948	Arabidopsis thalia
28	35	66.0	228	21	AAG43823	Arabidopsis thalia
29	35	66.0	228	21	AAG44223	Arabidopsis thalia
30	35	66.0	245	21	AAG40737	Zea mays protein f
31	35	66.0	245	21	AAG43822	Arabidopsis thalia
32	35	66.0	245	21	AAG44222	Arabidopsis thalia
33	35	66.0	285	21	AAG52356	Arabidopsis thalia
34	35	66.0	302	21	AAG52355	Arabidopsis thalia
35	35	66.0	303	21	AAG40736	Zea mays protein f
36	35	66.0	303	21	AAG43821	Arabidopsis thalia
37	35	66.0	303	21	AAG44221	Arabidopsis thalia
38	35	66.0	305	17	AAW00650	Protein phosphatas
39	35	66.0	305	22	AAM78980	Human protein SEQ
40	35	66.0	313	22	ABG05535	Novel human diagno
41	35	66.0	328	22	AAM79964	Human protein SEQ
42	35	66.0	[illegible]	[illegible]	AAM[illegible]	Human protein sequ
43	35	66.0	[illegible]	21	AAG52354	Arabidopsis thalia
44	35	66.0	[illegible]	21	AAG4[illegible]	Zea mays protein f
45	34	64.2	48	22	ABB37578	Peptide #5084 enco

ALIGNMENTS

```
RESULT  1
AAY82840
ID   AAY82840 standard; peptide; 9 AA.
XX
AC   AAY82840;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Lactadherin (BA-46) peptide fragment (tumour associated antigen).
XX
KW   Tumour associated antigen peptide, TAA, cancer, carcinoma;
KW   treatment; prevention; cure; anti tumour vaccine; metastases;
KW   breast, bladder, prostate, pancreas, ovary, thyroid, colon;
KW   stomach; carcinoma; MHC Class I; HLA-A2; human;
KW   Major Histocompatibility Complex; uroplakin;
KW   prostate specific antigen; prostate specific membrane antigen;
KW   prostate acid phosphatase, mucin, lactadherin;
KW   teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRIPTO-1.
XX
OS   Homo sapiens.
XX
PN   WO200006723-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-IL00417.
XX
PR   30-JUL-1998;   98IL-0125608.
XX
PA   (YEDA ) YEDA RES & DEV CO LTD.
PA   (BIOT ) BIO-TECHNOLOGY GEN CORP.
XX
PI   Eisenbach L,  Carmon L,  Tirosh B,  Bar haim E,  Paz A,  Fridkin M;
PI   Fitzer attas C,
```

```
XX
DR   WPI; 2000-205463/18.
XX
PT   Tumor associated antigen peptides, especially derived from uroplakin,
PT   useful as vaccines to prevent or cure cancers including breast,
PT   bladder, prostate, pancreas, ovary, thyroid, colon and stomach.
XX
PS   Claim 17; Page 99; 113pp; English.
XX
CC   Tumour associated antigen peptides (TAA) may be used for the
CC   treatment, prevention and cure of cancer or cancer metastases. The
CC   cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
CC   colon, stomach, head or neck cancer or a carcinoma. The tumour
CC   associated antigens are presentable to the immune system by HLA-A2
CC   molecules and are generally between 8 to 10 amino acids in length.
CC   The amino acids located at positions 2 and 9 of the tumour associated
CC   antigens are the anchor residues which participate in the binding to
CC   MHC class I molecules, more specifically HLA-A2.  More tumour
CC   associated antigens are described in GENESEQ records AAY82806-Y82882.
CC   Those tumour associated antigens described in records AAY82806-Y82824
CC   and AAY82855-YY82869 are derived from Uroplakin, such as Uroplakin II,
CC   Uroplakin Ia, Uroplakin III and Uroplakin Ib.  Those described in
CC   records AAY82825-Y82829 are derived from prostate specific antigen
CC   (PSA). Those described in records AAY82830-Y82835 are derived from
CC   prostate specific membrane antigen (PSMA).  Those described in
CC   records YY82836-AAY82839 are derived from prostate acid phosphotase
CC   (PAP).  Those described in records AAY82840-Y82846 are derived
CC   from Lactadherin (BA-46).  Those described in records AAY82847-Y82854
CC   are derived from Mucin and those described in records AAY82871-Y82882
CC   are derived from Teratocarcinoma derived growth factor (CRIPTO-1).
XX
SQ   Sequence   9 AA;


  Query Match          100.0%;  Score 53;  DB 21;  Length 9;
  Best Local Similarity  100.0%;  Pred. No. 6.4e-05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
           |||||||||
Db       1 kqgnfnawv 9


RESULT   2
AAR77253
ID   AAR77253 standard; Protein; 217 AA
XX
AC   AAR77253;
XX
DT   21-NOV-1995  (first entry)
XX
DE   HMFG 46 kDa antigen C-terminal region.
XX
KW   HMFG; human milk fat globule, antigen, immunogen; vaccine; virucide;
KW   epithelium; tumor; breast cancer; monoclonal antibody; MAb.
XX
OS   Homo sapiens.
XX
PN   WO9515171-A.
XX
PD   08-JUN-1995.
XX
PF   05-DEC-1994;   94WO-US13967.
XX
PR   03-DEC-1993;   93US-0162402
XX
PA   (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI   Ceriani RL,  Larocca DJ,  Peterson JA;
XX
DR   WPI; 1995-215151/28.
XX
PT   46 kD apparent molecular weight human milk fat globule antigen -
PT   used in assays to determine the presence of a cancerous tumour of
PT   epithelial origin, and in a vaccine against neoplastic tumours
XX
PS   Claim 7; Page 41; 68pp; English.
XX
CC   A partial cDNA clone BA46-1 (AAQ91199) for the 46 kDa HMFG antigen,
CC   a major component of the apical surface of the normal breast
CC   epithelial cell, was obtained by PCR of cDNA from a lactating
CC   breast cDNA library.  The C-terminal region of the encoded protein
CC   showed 43% identity to corresp. regions of human Factor-V and 38%
CC   to Factor-VIII.
XX
SQ   Sequence   217 AA;


  Query Match          100.0%;  Score 53;  DB 16;  Length 217;
  Best Local Similarity  100.0%;  Pred. No. 0.12;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
           |||||||||
Db     101 kqgnfnawv 109


RESULT   3
AAR77254
ID   AAR77254 standard; Protein; 218 AA.
XX
AC   AAR77254;
XX
DT   21-NOV-1995  (first entry)
XX
DE   HMFG 46 kDa antigen partial sequence.
XX
KW   HMFG; human milk fat globule, antigen, immunogen, vaccine; virucide;
KW   epithelium; tumor; breast cancer, monoclonal antibody; MAb.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   69..71
FT                   /label= N-glycosylation_site
FT   Modified-site   156..158
FT                   /label= N-glycosylation_site
FT   Modified-site   160..162
FT                   /label= N-glycosylation_site
FT   Modified-site   181..183
FT                   /label= N-glycosylation_site
XX
PN   WO9515171-A.
XX
PD   08-JUN-1995.
XX
PF   05-DEC-1994;   94WO-US13967.
XX
PR   03-DEC-1993;   93US-0162402
XX
PA   (CANC-) CANCER RES FUND CONTRA COSTA
XX
PI   Ceriani RL,  Larocca DJ,  Peterson JA;
XX
DR   WPI; 1995-215151/28.
DR   N-PSDB; AAQ91199.
XX
PT   46 kD apparent molecular weight human milk fat globule antigen -
PT   used in assays to determine the presence of a cancerous tumour of
PT   epithelial origin, and in a vaccine against neoplastic tumours
XX
PS   Example 7; Page 38-40; 68pp; English.
XX
CC   A partial cDNA clone BA46-1 (AAQ91199) for the 46 kDa HMFG antigen,
CC   a major component of the apical surface of the normal breast
CC   epithelial cell, was obtained by PCR of cDNA from a lactating
```

```
CC   breast cDNA library.
XX
SQ   Sequence   218 AA,


  Query Match             100.0%;  Score 53;  DB 16;  Length 218;
  Best Local Similarity   100.0%;  Pred. No. 0.12;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
           |||||||||
Db     102 kqgnfnawv 110


RESULT   4
AAR77252
ID   AAR77252 standard; Protein; 387 AA.
XX
AC   AAR77252;
XX
DT   21-NOV-1995  (first entry)
XX
DE   HMFG 46 kDa antigen.
XX
KW   HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
KW   epithelium; tumor; breast cancer; monoclonal antibody; MAb
XX
OS   Homo sapiens.
XX
PN   WO9515171-A.
XX
PD   08-JUN-1995.
XX
PF   05-DEC-1994;   94WO-US13967
XX
PR   03-DEC-1993;   93US-0162402.
XX
PA   (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI   Ceriani RL,  Larocca DJ,  Peterson JA;
XX
DR   WPI; 1995-215151/28
DR   N-PSDB; AAQ91198.
XX
PT   46 kD apparent molecular weight human milk fat globule antigen -
PT   used in assays to determine the presence of a cancerous tumour of
PT   epithelial origin, and in a vaccine against neoplastic tumours
XX
PS   Claim 6; Page 46-47; 68pp; English
XX
CC   A complete cDNA sequence for the 46 kDa HMFG antigen, a major
CC   component of the apical surface of the normal breast epithelial
CC   cell, was obtained by PCR and RACE methods.  cDNA clones can be
CC   used to prepare MAbs for use in immunotherapy, immunohistopathology,
CC   prognosis, diagnosis, imaging and therapy.  Recombinant antigen can
CC   be expressed in prokaryotic or (glycosylated) in eukaryotic cells.
XX
SQ   Sequence   387 AA;


  Query Match             100.0%;  Score 53;  DB 16;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
           |||||||||
Db     271 kqgnfnawv 279


RESULT   5
AAY94453
ID   AAY94453 standard; Protein; 387 AA.
XX
AC   AAY94453;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Human lactadherin protein.
XX
KW   Human; lactadherin; MGF-E8; anti-tumour; immune response;
KW   exosome; dendritic cell.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Secretion_signal
FT   Protein         24..387
FT                   /label= Lactadherin
FT   Binding-site    46..48
FT                   /label= Integrin_binding_site
XX
PN   EP1004664-A1.
XX
PD   31-MAY-2000.
XX
PF   24-NOV-1998;   98EP-0402925.
XX
PR   24-NOV-1998;   98EP-0402925.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (CURI-) INST CURIE.
XX
DR   WPI; 2000-352597/31.
DR   N-PSDB; AAA27140.
XX
PT   Chimeric isolated (human) lactadherin polypeptide that functions as an
PT   adaptor of cross-priming to eliminate pathogenic antigens, e.g. in
PT   tumors -
XX
PS   Example 3; Page 12; 20pp; English.
XX
CC   Lactadherin protein was found in exosomes produced by dendritic cells.
CC   The protein is involved in the phagocytosis of particulate antigens by
CC   dendritic cells. Exosomes produced by dendritic cells exposed to
CC   tumour antigens induce potent immune responses. Lactadherin or variants
CC   of it may be used in the mediation of an immune response. Variants of
CC   lactadherin may be used for inhibition and/or stimulation of the
CC   cross priming of antigens and stimulation of the phagocytosis of
CC   antigens by dendritic cells. Compositions derived from lactadherin can
CC   also be used to monitor an immune response, more specifically a CTL
CC   (cytotoxic T lymphocyte) response and also to produce CTLs specific
CC   for a selected antigen. The present sequence is the human
CC   lactadherin protein.
XX
SQ   Sequence   387 AA;


  Query Match             100.0%;  Score 53;  DB 21;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
           |||||||||
Db     271 kqgnfnawv 279


RESULT   6
AAB71682
ID   AAB71682 standard; protein; 363 AA.
XX
AC   AAB71682;
XX
DT   10-MAY-2001  (first entry)
XX
DE   Bos taurus interleukin-2 receptor gamma.
```

```
XX
KW   Cytokine receptor common gamma chain like; CRCGCL; human;
KW   tumours; infections; inflammatory; immune disorder;
KW   neurodegenerative; cardiovascular; disorder.
XX
OS   Bos taurus.
XX
PN   WO200112672 A2.
XX
PD   22-FEB-2001.
XX
PF   17-AUG-2000; 2000WO-US22493.
XX
PR   18-AUG-1999;   99US-0376440.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SA,  Rosen CA,  Moore PA;
XX
DR   WPI; 2001-147547/15.
XX
PT   New nucleic acid molecule encoding a human cytokine receptor common
PT   gamma chain like polypeptide, useful for treating, preventing and/or
PT   diagnosing e.g. tumors, inflammatory diseases and immunodeficiency
PT   conditions -
XX
PS   Disclosure; Fig 2; 288pp; English.
XX
CC   The present invention relates to a human cytokine receptor
CC   common gamma chain like protein (CRCGCL). The invention is useful
CC   for treating, preventing and/or diagnosing conditions such
CC   as tumours, infections, inflammatory diseases,
CC   immune disorders, neurodegenerative disorder and cardiovascular
CC   disorder.
XX
SQ   Sequence   363 AA;


  Query Match             73.6%;  Score 39;  DB 22;  Length 363;
  Best Local Similarity   85.7%;  Pred. No. 59;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 QGNFNAW 8
           ||||:||
Db     301 qgnfsaw 307


RESULT   7
AAR59094
ID   AAR59094 standard; Protein; 369 AA.
XX
AC   AAR59094;
XX
DT   04-MAY-1995  (first entry)
XX
DE   Murine IL-2R gamma.
XX
KW   Murine IL2-R gamma; X-linked severe combined immunodeficiency;
KW   XSCID; interleukin.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "signal peptide"
FT   Domain          258..284
FT                   /note= "transmembrane domain"
FT   Misc-difference 331
FT                   /note= "Coresponding codon CAG"
FT   Modified-site   71..73
FT                   /label= N-glycosylation_site
FT   Modified-site   75..77
FT                   /label= N-glycosylation_site
FT   Modified-site   84..86
FT                   /label= N-glycosylation_site
FT   Modified-site   96..98
FT                   /label= N-glycosylation_site
FT   Modified-site   159..161
FT                   /label= N-glycosylation_site
FT   Modified-site   255..257
FT                   /label= N-glycosylation_site
XX
PN   WO9420641-A.
XX
PD   15-SEP-1994.
XX
PF   10-MAR-1994;   94WO-US02891.
XX
PR   12-MAR-1993;   93US-0031143.
PR   14-SEP-1993;   93US-0121435.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Leonard WJ,  Mcbride WO,  Noguchi M;
XX
DR   WPI; 1994-303046/37.
DR   N-PSDB; AAQ71977.
XX
PT   Diagnosis of X-linked severe combined immunodeficiency (XSCID) -
PT   comprises detecting mutated IL 2R gamma gene, also vectors and
PT   transgenic animals containing the mutated gene
XX
PS   Example 1; Fig 7; 98pp; English.
XX
CC   AAQ71977 is the DNA sequence of murine IL-2R gamma AAR59094,
CC   this was used in the development of a claimed method for the
CC   diagnosis of X linked severe combined immunodeficiency (XSCID),
CC   in female carriers and male sufferers.
XX
SQ   Sequence   369 AA;


  Query Match             73.6%;  Score 39;  DB 15;  Length 369;
  Best Local Similarity   85.7%;  Pred. No. 60;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 QGNFNAW 8
           ||||:||
Db     304 qgnfsaw 310


RESULT   8
ABG29550
ID   ABG29550 standard; Protein; 673 AA
XX
AC   ABG29550;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #29541.
XX
KW   Human, chromosome mapping, gene mapping, gene therapy, forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067 A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
```

```
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS93737.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 59909; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   673 AA;


  Query Match             71.7%;  Score 38;  DB 22;  Length 673;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 GNFNAWV 9
           |||:|||
Db     432 gnfsawv 438


RESULT  9
ABG24996
ID   ABG24996 standard; Protein; 1087 AA.
XX
AC   ABG24996;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #24987.
XX
KW   Human; chromosome mapping; gene mapping, gene therapy, forensic,
KW   food supplement, medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0644167
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS89183.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 55355; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1087 AA;


  Query Match             71.7%;  Score 38;  DB 22;  Length 1087;
  Best Local Similarity   85.7%;  Pred. No. 2.8e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 GNFNAWV 9
           |||:|||
Db     846 gnfsawv 852


RESULT  10
AAM65565
ID   AAM65565 standard; Protein; 51 AA.
XX
AC   AAM65565;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe encoded protein SEQ ID NO: 87670.
XX
KW   Human; brain expressed exon, gene expression analysis; probe;
KW   microarray, Alzheimer's disease, multiple sclerosis, Schizophrenia;
KW   epilepsy; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157275 A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   25-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
```

PR 27-SEP-2000; 2000US-0236259.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR,
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO 37670; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention
XX
SQ Sequence 51 AA;

Query Match 67.6%; Score 36; DB 22; Length 51;
Best Local Similarity 75.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KQGNFNAW 8
|||: |||
Db 48 kqgkvnaw 45

RESULT 11
AAB65971
ID AAB65971 standard; protein; 68 AA.
XX
AC AAB65971;
XX
DT 03-APR-2001 (first entry)
XX
DE Human secreted protein BLAST search protein SEQ ID NO. 111.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cytostatic; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PN WO200077023-A1.
XX
PD 21-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US14964.
XX
PR 11-JUN-1999; 99US-0138627.
XX
PA (HUMA-) HUMAN GENOME SCI INC
XX
PI Rosen CA, Ruben SM, Komatsoulis GA;
XX
DR WPI; 2001-032308/04.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; Page 511; 558pp; English.
XX
CC The invention relates to the isolation of genes AAB45071-B45120 encoding
CC the human secreted proteins AAB65920-B65968. This sequence represents a
CC peptide fragment homologous to the protein encoded by the gene given
CC in the descriptor line. The sequence is a search result from a BLASTX
CC homology search. The genes and proteins are useful for preventing,
CC ameliorating or treating medical conditions, e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such
CC as myocardial ischaemias; (d) wound healing; (e) neurological diseases
CC e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such
CC as viral, bacterial, fungal and parasitic infections.
XX
SQ Sequence 68 AA;

Query Match 67.9%; Score 36; DB 22; Length 68;
Best Local Similarity 55.6%; Pred. No. 33;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KQGNFNAWV 9
:: |:|||
Db 10 eknnynawv 18

RESULT 12
AAB04068
ID AAB04068 standard; Protein; 113 AA.
XX
AC AAB04068;
XX
DT 11-APR-2001 (first entry)
XX
DE Rat zcytor 10 cytokine receptor peptide fragment.
XX
KW zcytor 10 cytokine receptor; cytokine; receptor; antibody; ligand;
KW binding; detection; modulation; recombinant cell;
KW haematopoietic cell, lymphoid cell, myeloid cell; lymph;
KW immune system, blood, bone, inflammatory response; inflammation;
KW spleen; rat.
XX
OS Rattus rattus.
XX
PN WO200068381-A1.
XX
PD 16-NOV-2000.
XX
PF 11-MAY-2000; 2000WO-US12924.
XX
PR 11-MAY-1999; 99US-0309861.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Presnell SR, Foster DC, Hammond AK, Lok S;
XX
DR WPI; 2001-016096/02.
DR N-PSDB; AAA54463.
XX
PT New cytokine receptor mouse zcytor 10, useful for detecting ligands
PT that stimulate proliferation or development of haematopoietic,
PT lymphoid and myeloid cells
XX
PS Example 3; Page 115; 134pp; English.
XX
CC Isolating a nucleotide which encodes the zcytor 10 cytokine
CC receptor enables the production of recombinant cells expressing the
CC receptor. Those cells can then be used to detect the presence of a
CC modulator of zcytor10 protein by culturing the cells in the presence
CC of a test ligand and comparing levels of activity of mouse zcytor10

```
CC   in the presence and absence of the test sample. Similarly, detection
CC   of zcytor10 receptor ligand within a test sample can be achieved.
CC   The method comprising contacting a test sample containing an amino
CC   acid sequence from Cys15 to Gly27 to ... of the zcytor 10
CC   cytokine receptor and detecting the binding of the polypeptide to a
CC   ligand in the sample. Specified peptide fragments of the zcytor 10
CC   cytokine receptor and the methods described are used to identify
CC   ligands that stimulate the proliferation and/or development of
CC   haematopoietic, lymphoid and myeloid cells. Peptide fragments of
CC   the cytokine receptor are useful for treating lymphoid, immune,
CC   inflammatory, splenic, blood or bone disorders and for generating
CC   antibodies directed against the receptor.
XX
SQ   Sequence   113 AA;

  Query Match             67.9%;  Score 36;  DB 22;  Length 113;
  Best Local Similarity   71.4%;  Pred. No. 57;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         3 GNFNAWV 9
             ||| ||:
Db        41 gniqawi 47

RESULT 13
AAB71691
ID   AAB71691 standard; peptide; 181 AA.
XX
AC   AAB71691;
XX
DT   10-MAY-2001  (first entry)
XX
DE   CRCGCL related peptide #4.
XX
KW   Cytokine receptor common gamma chain like; CRCGCL; human;
KW   tumours; infections; inflammatory; immune disorder;
KW   neurodegenerative; cardiovascular; disorder.
XX
OS   Homo sapiens.
XX
PN   WO200112672-A2.
XX
PD   22-FEB-2001.
XX
PF   17-AUG-2000; 2000WO-US22493.
XX
PR   18-AUG-1999;   99US-0376430.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SA,  Rosen CA,  Moore PA;
XX
DR   WPI; 2001-147547/15.
XX
PT   New nucleic acid molecule encoding a human cytokine receptor common
PT   gamma chain like polypeptide, useful for treating, preventing and/or
PT   diagnosing e.g. tumors, inflammatory diseases and immunodeficiency
PT   conditions -
XX
PS   Disclosure; Page 18; 288pp; English.
XX
CC   The present invention relates to a human cytokine receptor
CC   common gamma chain like protein (CRCGCL). The invention is useful
CC   for treating, preventing and/or diagnosing conditions such
CC   as tumours, infections, inflammatory diseases,
CC   immune disorders, neurodegenerative disorder and cardiovascular
CC   disorder.
XX
SQ   Sequence   181 AA;

  Query Match             67.9%;  Score 36;  DB 22;  Length 181;
  Best Local Similarity   62.5%;  Pred. No. 94;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         2 QGNFNAWV 9
              |||| |:
Db        91 qgnfqewi 98

RESULT 14
AAW94697
ID   AAW94697 standard; Protein; 321 AA.
XX
AC   AAW94697;
XX
DT   05-MAY-1999  (first entry)
XX
DE   Human milk fat globule protein MFG-E8.
XX
KW   Del 1, developmentally regulated endothelial cell locus 1; cancer;
KW   discoidin I; factor VIII-like domain; epidermal growth factor; EGF;
KW   diabetic retinopathy; rheumatoid arthritis; endometriosis;
KW   angiogenesis.
XX
OS   Homo sapiens.
XX
PN   US5877281-A.
XX
PD   02-MAR-1999.
XX
PF   05-JUN-1996;   96US-0659235.
XX
PR   05-JUN-1996;   96US-0659235.
PR   07-JUN-1995;   95US-0480229.
XX
PA   (PROG-) PROGENITOR INC.
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Hogan B,  Quertermous T,  Snodgrass HR,  Zupancic TJ;
XX
DR   WPI; 1999-189720/16.
XX
PT   Del 1 polypeptide sequences - useful for treatment of cancer,
PT   diabetic retinopathy, rheumatoid arthritis and endometriosis
XX
PS   Example, Column 63 64, 73pp, English.
XX
CC   The present sequence represents human milk fat globule protein MFG-E8,
CC   which has homology to the developmentally regulated endothelial cell
CC   locus 1 (Del 1). The Del-1 protein has epidermal growth factor (EGF) like
CC   domains and discoidin I/factor VIII-like domains. The Del-1 proteins
CC   have an inhibitory effect on angiogenesis (blood vessel growth), this
CC   activity may be useful clinically to prevent neovascularisation of
CC   tissues such as tumour nodules and prevention of metastases. The anti-
CC   angiogenic activity of Del-1 may be used to treat abnormal conditions
CC   that result from angiogenesis, including cancer, diabetic retinopathy,
CC   rheumatoid arthritis and endometriosis. Since Del-1 promotes
CC   angiogenesis it can be used to treat cardiac ischaemia, thrombotic
CC   stroke, wound healing and peripheral vascular disease. Del 1 is also
CC   useful for promoting bone formation. Del-1 binds to alpha V beta 3
CC   integrin, and is an apoptosis inducer.
XX
SQ   Sequence   321 AA;

  Query Match             67.9%;  Score 36;  DB 20;  Length 321;
  Best Local Similarity   75.0%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         1 KQGNFNAW 8
             |||  |||
Db       204 kqgkvnaw 211
```

```
RESULT 15
AAB04066
ID   AAB04066 standard; Protein; 357 AA.
XX
AC   AAB04066;
XX
DT   11-APR-2001 (first entry)
XX
DE   Zcytor 10 cytokine receptor.
XX
KW   zcytor 10 cytokine receptor, cytokine, receptor, antibody; ligand;
KW   binding; detection; modulation; recombinant cell;
KW   haematopoietic cell; lymphoid cell; myeloid cell; lymph;
KW   immune system; blood; bone; inflammatory response; inflammation;
KW   spleen; human.
XX
OS   Homo sapiens.
XX
PN   WO200068381-A1.
XX
PD   16-NOV-2000.
XX
PF   11-MAY-2000; 2000WO-US12924.
XX
PR   11-MAY-1999;   99US-0309861.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Presnell SR, Foster DC, Hammond AK, Lok S;
XX
DR   WPI; 2001-016096/02.
DR   N-PSDB; AAA54451.
XX
PT   New cytokine receptor mouse zcytor 10, useful for detecting ligands
PT   that stimulate proliferation or development of haematopoietic,
PT   lymphoid and myeloid cells
XX
PS   Claim 1; Page 109-110; 134pp; English.
XX
CC   Isolating a nucleotide which encodes the zcytor 10 cytokine
CC   receptor enables the production of recombinant cells expressing the
CC   receptor. Those cells can then be used to detect the presence of a
CC   modulator of zcytor10 protein by culturing the cells in the presence
CC   of a test ligand and comparing levels of activity of mouse zcytor10
CC   in the presence and absence of the test sample. Similarly, detection
CC   of zcytor10 receptor ligand within a test sample can be achieved.
CC   The method comprising contacting a test sample containing an amino
CC   acid sequence from Cys15 or Gly25 to Pro230 of the zcytor 10
CC   cytokine receptor and detecting the binding of the polypeptide to a
CC   ligand in the sample. Specified peptide fragments of the zcytor 10
CC   cytokine receptor and the methods described are used to identify
CC   ligands that stimulate the proliferation and/or development of
CC   haematopoietic, lymphoid and myeloid cells. Peptide fragments of
CC   the cytokine receptor are useful for treating lymphoid, immune,
CC   inflammatory, splenic, blood or bone disorders and for generating
CC   antibodies directed against the receptor.
XX
SQ   Sequence   357 AA;


  Query Match             67.9%;  Score 36;  DB 22;  Length 357;
  Best Local Similarity   71.4%;  Pred. No. 1.9e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CNFNAWV 9
           ||| ||:
Db     280 qnfqawi 286



Search completed: September  5, 2002, 15:26:28
Job time:  468 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:22:55 ; Search time 60.78 Seconds
(without alignments)
12.393 Million cell updates/sec

Title: US-09-744-804-35
Perfect score: 53
Sequence: 1 KQGNFNAWV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	53	100.0	218	2	A47285	milk fat globule p
2	42	79.2	734	2	S00704	photosystem I prot
3	42	79.2	734	2	T07544	photosystem I prot
4	40	75.5	268	2	H90826	hypothetical prote
5	40	75.5	386	2	H85684	unknown protein en
6	39	73.6	127	2	AB0154	probable bacteriop
7	39	73.6	369	2	I49280	interleukin 2 rece
8	39	73.6	734	1	A2NTP7	photosystem I P700
9	39	73.6	734	2	S00445	photosystem I prot
10	39	73.6	734	2	S60184	photosystem I prot
11	38	71.7	61	2	C90821	hypothetical prote
12	38	71.7	64	2	D90832	hypothetical prote
13	38	71.7	64	2	F85689	unknown protein en
14	38	71.7	74	2	E90901	hypothetical prote
15	38	71.7	74	2	H85743	unknown protein en
16	38	71.7	401	2	S65138	glycoprotein antig
17	38	71.7	409	2	T11743	pP47 protein - pig
18	38	71.7	427	2	S74211	PAS-6/7 protein pr
19	38	71.7	734	1	A2R2P7	photosystem I P700
20	38	71.7	734	1	A2LVP7	photosystem I P700
21	38	71.7	735	2	S50551	photosystem I prot
22	37	69.8	373	2	AI0440	probable aliphatic
23	37	69.8	502	2	S64860	hypothetical prote
24	37	69.8	733	2	S78376	photosystem I P700
25	37	69.8	734	2	T07280	photosystem I P700
26	37	69.8	735	2	S41482	P700 chlorophyll a
27	37	69.8	736	2	B28341	photosystem I P700
28	37	69.8	737	2	T06839	probable photosyst
29	37	69.8	741	1	S20923	photosystem I prot
30	37	69.8	817	2	S77106	hypothetical prote
31	36	67.9	278	2	A60195	transforming prote
32	36	67.9	359	2	JC7280	cytokine receptor-
33	36	67.9	434	2	AG0071	ExsI transport pro
34	36	67.9	472	1	B65098	hexuronate transpo
35	36	67.9	472	2	C91125	transport protein
36	36	67.9	472	2	F85970	transport of hexur
37	36	67.9	503	2	S33912	aromatase (EC 1.14
38	36	67.9	508	1	A36121	aromatase (EC 1.14
39	36	67.9	526	2	C84790	hypothetical prote
40	36	67.9	731	2	S18243	photosystem I prot
41	36	67.9	733	2	S06902	photosystem I prot
42	36	67.9	734	1	S26072	photosystem I prot
43	36	67.9	734	2	S77206	photosystem I prot
44	36	67.9	741	2	I39616	photosystem I prot
45	36	67.9	741	2	AC2452	photosystem I core

ALIGNMENTS

RESULT 1
A47285
milk fat globule protein human (fragment)
C;Species: Homo sapiens (man)
C;Date: 31-Dec 1992 #sequence_revision 31-Dec 1992 #text_change 29 May 1996
C;Accession: A47285
R;Larocca, D.; Peterson, J.A.; Urrea, R.; Kuniyoshi, J.; Bistrain, A.M.; Ceriani, R.L.
Cancer Res. 51, 4994-4998, 1991
A;Title: A MCF-7,200 human milk fat globule protein that is highly expressed in huma
A;Reference number: A47285; MUID:91371351
A;Accession: A47285
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-218 <LAR>
A;Cross-references: GB:S56161; NID:g456496; PIDN:AAB19771.1; PID:g235397
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; FGF hom
F;1-56/Domain: discoidin I amino terminal homology (fragment) <DN1>
F;60-218/Domain: discoidin I amino terminal homology <DN2>

Query Match 100.0%; Score 53; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KQGNFNAWV 9
||||||||| |
Db 102 KQGNFNAWV 110

RESULT 2
S00704
photosystem I protein A2 - garden pea chloroplast
C;Species: chloroplast Pisum sativum (garden pea)
C;Date: 28-Aug 1989 #sequence_revision 28-Aug 1989 #text_change 26-Aug-1999
C;Accession: S00704; R30830
R;Lehmbeck, J.; Rasmussen, O.F.; Bookjans, G.B.; Jepsen, B.R.; Stummann, B.M.; Hennin
Plant Mol. Biol. 7, 3-10, 1986
A;Title: Sequence of two genes in pea chloroplast DNA coding for 84 and 82 kD polypep
A;Reference number: S00703
A;Accession: S00704
A;Molecule type: DNA
A;Residues: 1-734 <LEH>
A;Cross-references: EMBL:X05423; NID:g22134; PIDN:CAA28901.1; PID:g22136
A;Note: the authors translated the codon AAT for residue 115 as Ile and ATT for resid
C;Genetics:
A;Gene: psaA2
A;Genome: chloroplast
C;Superfamily: photosystem I P700 apoprotein
C;Keywords: chloroplast; electron transfer; membrane-associated complex; photosynthes

Query Match 79.2%; Score 42; DB 2; Length 734;

```
  Best Local Similarity   87.5%;  Pred. No. 6.9;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
            |||| |||
Db      71 QGNFEAWV 78


RESULT   3
T07544
photosystem I protein A2 - Japanese black pine chloroplast
C;Species: chloroplast Pinus thunbergiana (Japanese black pine)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 18-Aug-2000
C;Accession: T07544
R;Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugiura, M.
Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994
A;Title: Loss of all ndh genes as determined by sequencing the entire chloroplast genome
A;Reference number: Z16030; MUID:95024047
A;Accession: T07544
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-734 <WAK>
A;Cross-references: EMBL:D17510; NID:g529643; PIDN:BAA04420.1; PID:g1262705
C;Genetics:
A;Gene: psaB
A;Genome: chloroplast
C;Superfamily: photosystem I P700 apoprotein
C;Keywords: chloroplast; membrane associated complex; photosystem I


  Query Match             79.2%;  Score 42;  DB 2;  Length 734;
  Best Local Similarity   87.5%;  Pred. No. 6.9;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
            |||| |||
Db      71 QGNFEAWV 78


RESULT   4
H90826
hypothetical protein ECs1584 [imported] - Escherichia coli (strain O157:H7, substrain R
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: H90826
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A90629; MUID:21156231; PMID:11258796
A;Accession: H90826
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-268 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB35007.1; PID:g13361048; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1584


  Query Match             75.5%;  Score 40;  DB 2;  Length 268;
  Best Local Similarity   87.5%;  Pred. No. 5.9;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
            | |||||
Db     135 QHNFNAWV 142


RESULT   5
H85684
unknown protein encoded by prophage CP-933P [imported] - Escherichia coli (strain O157:H
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: H85684
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, E.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.T.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: H85684
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-386 <PER>
A;Cross-references: GB:AF005174; NID:g12514762; PIDN:AAG55940.1; GSPDB:GN00115; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z1843


  Query Match             75.5%;  Score 40;  DB 2;  Length 386;
  Best Local Similarity   87.5%;  Pred. No. 8.5;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
            | ||||||
Db     253 QHNFNAWV 260


RESULT   6
AB0154
probable bacteriophage protein (partial) YPO1259 [imported] - Yersinia pestis (strain
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AB0154
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AB0154
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-127 <KUF>
A;Cross-references: GB:AL590842; PIDN:CAC92093.1; PID:g15979313; GSPDB:GN00175
C;Genetics:
A;Gene: YPO1259


  Query Match             73.6%;  Score 39;  DB 2;  Length 127;
  Best Local Similarity   75.0%;  Pred. No. 4.3;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
            | |||||:
Db      19 QHNFNAWI 26


RESULT   7
I49280
interleukin-2 receptor gamma chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 20-Jan-2000
C;Accession: I49280; A47514; JN0592; JN0275; S37582; I53398
R;Cao, X.; Kozak, C.A.; Liu, Y.
Proc. Natl. Acad. Sci. U.S.A. 90, 8464-8468, 1993
A;Title: Characterization of cDNAs encoding the murine interleukin 2 receptor (IL-2R)
A;Reference number: A47514; MUID:93391374
A;Accession: I49280
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-369 <CAO>
A;Cross-references: EMBL:U21795; NID:g727349; PIDN:AAA64279.1; PID:g727350
A;Accession: A47514
A;Status: translated from GB/EMBL/DDBJ
```

A;Molecule type: mRNA
A;Residues: 1-369 <RE2>
A;Cross-references: GB:L20048; NID:g404067; PIDN:AAA39285.1; PID:g404068
F.Kumaki, S.; Kondo, M.; Takeshita, T.; Asao, H.; Nakamura, M.; Sugamura, K.
Biochem. Biophys. Res. Commun. 193, 356-363, 1993
A;Title: Cloning of the mouse interleukin 2 receptor gamma chain: Demonstration of funct
A;Reference number: JN0592; MUID:93277575
A;Accession: JN0592
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-369 <KUM>
A;Cross-references: [redacted]; NID:g306684; PIDN:BAA02766.1; PID:g306685
R;Kobayashi, N.; Nakagawa, S.; Minami, Y.; Taniguchi, T.; Kono, T.
Gene 130, 303-304, 1993
A;Title: Cloning and sequencing of the cDNA encoding a mouse IL-2 receptor gamma.
A;Reference number: JN0775; MUID:93366191
A;Accession: JN0775
A;Molecule type: mRNA
A;Residues: 1-369 <KOG>
A;Cross-references: GB:D13821; NID:g436945; PIDN:BAA02974.1; PID:g436946
R;Chiu, R.K.; Dougherty, G.J.
submitted to the EMBL Data Library, October 1993
A;Description: Regulation of CD44-mediated cellular adhesion by the IL-2 R gamma chain.
A;Reference number: S37582
A;Accession: S37582
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-350,'S',352-366,'S',368-369 <CHI>
A;Cross-references: EMBL:X75337
R;DiSanto, J.P.; Certain, S.; Wilson, A.; MacDonald, H.R.; Avner, P.; Fischer, A.; de Sa
Eur. J. Immunol. 24, 3014-3018, 1994
A;Title: The murine interleukin 2 receptor gamma chain gene: organization, chromosomal l
A;Reference number: I53398; MUID:95104285
A;Accession: I53398
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-369 <RES>
A;Cross-references: GB:S75852; NID:g861554; PIDN:AAB32904.1; PID:g861555
C;Genetics:
A;Gene: IL-2Rgamma
A;Introns: 39/1; 90/2; 151/1; 199/3; 254/1; 290/2; 338/3
C;Complex: The high affinity receptor is a heterotrimer of alpha (see I15.UHMS2), beta (
eptors.
C;Function:
A;Description: receptor for interleukin 2
A;Pathway: interleukin 2 stimulated growth and differentiation of T cells, B cells, NK c
C;Superfamily: interleukin-2 receptor gamma chain
C;Keywords: cytokine receptor; duplication; glycoprotein; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-369/Product: interleukin-2 receptor gamma chain #status predicted <MAT>
F;256-284/Domain: transmembrane #status predicted <TMM>
F;71,75,84,96,159,164,306/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 73.6%; Score 39; DB 2; Length 369;
Best Local Similarity 85.7%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QGNFNAW 8
||||:|| |
Db 304 QGNFSAW 310

RESULT 8
A2NTP7
photosystem I P700 apoprotein A2 - common tobacco, chloroplast
C;Species: chloroplast Nicotiana tabacum (common tobacco)
C;Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 17-Feb-1995
C;Accession: A03466
R;Sugiura, M.
submitted to the EMBL Data Library, August 1986
A;Reference number: A00149
A;Accession: A03466
A;Molecule type: DNA
A;Residues: 1-734 <SUG>
A;Experimental source: cv. Bright Yellow 4
F.Shinozaki, K.; Ohme, M.; Tanaka, M.; Wakasugi, T.; Hayashida, N.; Matsubayashi, T.; Zs
Ohno, H.; Kamogasaira, T.; Yamada, K.; Kusuda, J.; Takaiwa, F.; Kato, A.; Tohdoh, N.;
EMBO J. 5, 2043-2049, 1986
A;Title: The complete nucleotide sequence of the tobacco chloroplast genome: its gene
A;Reference number: A38013
A;Contents: annotation; gene organization; sites, features
C;Comment: This is one of the specific proteins associated with chlorophyll a of the
s supplied by photosystem II to 2 substance (an iron sulfur protein), which in turn r
llar proteins.
C;Genetics:
A;Gene: psaB
A;Genome: chloroplast
C;Superfamily: photosystem I P700 apoprotein
C;Keywords: chloroplast; electron transfer; membrane protein; photosynthesis; photosy

Query Match 73.6%; Score 39; DB 1; Length 734;
Best Local Similarity 75.0%; Pred. No. 25;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 QGNFNAWV 9
|||| :||
Db 71 QGNFESWV 78

RESULT 9
S00445
photosystem I protein A2 - spinach chloroplast
C;Species: chloroplast Spinacia oleracea (spinach)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 20-Aug-1999
C;Accession: S00445
A;Kirsch, W., Seyet, P.; Herrmann, R.G.
Curr. Genet. 10, 843-855, 1986
A;Title: Nucleotide sequence of the clustered genes for two P700 chlorophyll a apopro
A;Reference number: S00444
A;Accession: S00445
A;Molecule type: DNA
A;Residues: 1-734 <KIR>
A;Cross-references: EMBL:X04121; NID:g12675; PIDN:CAA27745.1; PID:g12871
A;Note: the authors translated in fig.1 the codon ACT for residue 61 as Tyr
C;Genetics:
A;Gene: psaB
A;Genome: chloroplast
C;Superfamily: photosystem I P700 apoprotein
C;Keywords: chloroplast; electron transfer; membrane associated complex; photosynthes

Query Match 73.6%; Score 39; DB 2; Length 734;
Best Local Similarity 75.0%; Pred. No. 25;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 QGNFNAWV 9
|||| :||
Db 71 QGNFESWV 78

RESULT 10
S60184
photosystem I protein A2 - garden snapdragon chloroplast
N;Alternate names: photosystem I P700 apoprotein A2; photosystem I reaction center pr
C;Species: chloroplast Antirrhinum majus (garden snapdragon)
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 26-Aug-1999
C;Accession: S60184, S02395, S02394
R;Schaffner, C.; Laasch, H.; Hagemann, R.
Mol. Gen. Genet. 249, 533-544, 1995
A;Title: Detection of point mutations in chloroplast genes of Antirrhinum majus L. I.
A;Reference number: S60184; MUID:96133694
A;Accession: S60184
A;Molecule type: DNA
A;Residues: 1-734 <SCH>

```
A;Cross-references: EMBL:X84154
R;Schaffner, C.; Lausch, H.; Hagemann, R.
submitted to the EMBL Data Library, January 1995
A;Description: Detection of point mutations in chloroplast genes of Antirrhinum majus L.
A;Reference number: S52394
A;Accession: S52395
A;Molecule type: DNA
A;Residues: 1-566,'T',568-734 <SCW>
A;Cross-references: EMBL:X84153, NID:g669095, PIDN:CAA58958.1, PID:g669096
A;Accession: S52394
A;Molecule type: DNA
A;Residues: 1-135 <SC2>
A;Cross-references: EMBL:X84152; NID:g669093; PIDN:CAA58957.1; PID:g669094
C;Genetics:
A;Gene: psaB
A;Genome: chloroplast
C;Superfamily: photosystem I P700 apoprotein
C;Keywords: chloroplast; electron transfer; membrane-associated complex; photosynthesis;


  Query Match             73.6%;  Score 39;  DB 2;  Length 734;
  Best Local Similarity   75.0%;  Pred. No. 25;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
           |||| :||
Db      71 QGNFESWV 78


RESULT  11
C90821
hypothetical protein ECs1539 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: C90821
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
qasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C90821
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-61 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34962.1; PID:g13361003; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1539


  Query Match             71.7%;  Score 38;  DB 2;  Length 61;
  Best Local Similarity   85.7%;  Pred. No. 3.2;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 GNFNAWV 9
           |||:|||
Db      35 GNFSAWV 41


RESULT  12
D90832
hypothetical protein ECs1628 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: D90832
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
qasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90832
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-64 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB35051.1; PID:g13361092; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1628


  Query Match             71.7%;  Score 38;  DB 2;  Length 64;
  Best Local Similarity   85.7%;  Pred. No. 3.4;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 GNFNAWV 9
           |||:|||
Db      35 GNFSAWV 41


RESULT  13
F85689
unknown protein encoded by prophage CP-933X [imported] - Escherichia coli (strain O15
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: F85689
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: F85689
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-64 <STO>
A;Cross-references: GB:AE005174; NID:g12514809; PIDN:AAG55478.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z1881


  Query Match             71.7%;  Score 38;  DB 2;  Length 64;
  Best Local Similarity   85.7%;  Pred. No. 3.4;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 GNFNAWV 9
           |||:|||
Db      35 GNFSAWV 41


RESULT  14
E90901
hypothetical protein ECs2181 [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: E90901
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
qasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and q
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E90901
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-74 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB35604.1; PID:g13361647; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs2181


  Query Match             71.7%;  Score 38;  DB 2;  Length 74;
  Best Local Similarity   85.7%;  Pred. No. 3.9;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 GNFNAWV 9
           |||:|||
```

```
Db      35 GNFSAWV 41


RESULT  15
H85743
unknown protein encoded within prophage CP 933R [imported] - Escherichia coli (strain O1
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: H85743
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K., Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: H85743
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-74 <STO>
A;Cross-references: GB:AF005174; NID:g12515361; PIDN:AAG56412.1; GSPDB:GN00145; UWGP:Z23
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z2366


  Query Match             71.7%;  Score 38;  DB 2;  Length 74;
  Best Local Similarity   85.7%;  Pred. No. 3.9;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 GNFNAWV 9
           |||:|||
Db      35 GNFSAWV 41



Search completed: September  5, 2002, 15:28:49
Job time: 354 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:27:45 ; Search time 33.99 Seconds
(without alignments)
10.252 Million cell updates/sec

Title: US-09-744-804-35
Perfect score: 53
Sequence: 1 KQGNFNAWV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	53	100.0	387	1	MFGM_HUMAN	Q08431 homo sapien
2	42	79.2	734	1	PSAB_LOTJA	P58385 lotus japon
3	42	79.2	734	1	PSAB_OENHO	Q9mtn7 oenothera h
4	42	79.2	734	1	PSAB_PEA	P05311 pisum sativ
5	42	79.2	734	1	PSAB_PINTH	P41640 pinus thunb
6	41	77.4	734	1	PSAB_CYACA	Q9tlq6 cyanidium c
7	39	73.6	369	1	CYRG_MOUSE	P34902 mus musculu
8	39	73.6	379	1	CYRG_BOVIN	Q95118 bos taurus
9	39	73.6	734	1	PSAB_ANTMA	Q33332 antirrhinum
10	39	73.6	734	1	PSAB_SPIOL	P06512 spinacia ol
11	39	73.6	734	1	PSAB_TOPAC	P06407 nicotiana t
12	38	71.7	409	1	MFGM_PIG	P79385 sus scrofa
13	38	71.7	427	1	MFGM_BOVIN	Q95114 bos taurus
14	38	71.7	734	1	PSAB_ARATH	P56767 arabidopsis
15	38	71.7	734	1	PSAB_MARPO	P06408 marchantia
16	38	71.7	734	1	PSAB_ORYSA	P12156 oryza sativ
17	38	71.7	734	1	PSAB_WHEAT	P58386 triticum ae
18	38	71.7	735	1	PSAB_MAIZE	P04967 zea mays (m
19	38	71.7	994	1	BGAL_STAXY	Q83815 staphylococ
20	37	69.8	733	1	PSAB_ODOSI	P49480 odontella s
21	37	69.8	734	1	PSAB_CHLVU	P56342 chlorella v
22	37	69.8	734	1	PSAB_GUITH	Q78507 guillardia
23	37	69.8	734	1	PSAB_MESVI	Q9mur7 mesostigma
24	37	69.8	735	1	PSAB_CHLMO	P56492 chlamydomon
25	37	69.8	735	1	PSAB_CHLRE	P09144 chlamydomon
26	37	69.8	737	1	PSAB_CYAPA	P48113 cyanophora
27	37	69.8	738	1	PSAB_SYNPW	Q9r6t9 synechococc
28	37	69.8	740	1	PSAB_SYNPN	P25897 synechococc
29	37	69.8	747	1	PSAB_PROMA	Q9rdv0 prochlorocs
30	36	67.9	473	1	EXPI_ECOLI	P42609 escherichia
31	36	67.9	503	1	CPV1_MOUSE	P28649 mus musculu
32	36	67.9	508	1	CPV1_RAT	P22443 rattus norv
33	36	67.9	730	1	PSAB_SYNY3	P29255 synechocyst
34	36	67.9	732	1	PSAB_SYNE2	P17155 synechococc
35	36	67.9	734	1	PSAB_EUGGR	P19431 euglena gra
36	36	67.9	734	1	PSAB_PORPU	P51289 porphyra pu
37	36	67.9	740	1	PSAB_ANASP	P58565 anabaena sp
38	36	67.9	740	1	PSAB_ANAVA	P41088 anabaena va
39	36	67.9	742	1	PSAB_MASLA	Q52475 mastigoclad
40	36	67.9	1384	1	CTA1_HUMAN	P78357 homo sapien
41	35	66.0	147	1	SPAM_SALTY	P19912 salmonella
42	35	66.0	284	1	RORL_ERYSP	P26280 erythrobact
43	35	66.0	305	1	PPP6_HUMAN	O00743 homo sapien
44	35	66.0	305	1	PPP6_RAT	Q64620 rattus norv
45	35	66.0	626	1	PEPO_LACLA	Q07744 lactococcus

ALIGNMENTS

RESULT 1
MFGM_HUMAN
ID MFGM_HUMAN STANDARD; PRT; 387 AA.
AC Q08431;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMFG)
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
GN MFGE8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast, and Breast carcinoma;
RX MEDLINE=96213908; PubMed=8639264;
RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
RT "Cloning and sequence analysis of human breast epithelial antigen
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal
RT growth factor-like domain.";
RL DNA Cell Biol. 15:281-286(1996).
RN [2]
RP SEQUENCE OF 170-387 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=91371251; PubMed=1909942;
RA Tatoud R., Peterson J.A., Ortiz F., Kuniyoshi J., Bistrain A.M.,
RA Ceriani R.L.;
RT "A Mr 46,000 human milk fat globule protein that is highly expressed
RT in human breast tumors contains factor VIII-like domains.";
RL Cancer Res. 51:4994-4998(1991).
RN [3]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=98194924; PubMed=9535276;
RA Giuffrida M.G., Cavaletto M., Giunta C., Conti A.,
RA Godovac-Zimmermann J.;
RT "Isolation and characterization of full and truncated forms of human
RT breast carcinoma protein BA46 from human milk fat globule membranes.";
RL J. Protein Chem. 17:143-148(1998).
RN [4]
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
RX MEDLINE=99342076; PubMed=10411943;
RA Haeggstroem B., Naeslund J., Sletten K., Westermark G.T., Mucchiano G.,
RA Tjernberg L.O., Nordstedt C., Engstroem U., Westermark P.;
RT "Medin, an integral fragment of aortic smooth muscle cell-produced
RT lactadherin forms the most common human amyloid.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=97455895; PubMed=9260929;
RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
RT (RGD)-dependent cell adhesion.";

```
RL   DNA Cell Biol. 16:861-869(1997).
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
CC       SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
CC   -!- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
CC       MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC   -!- PTM: MEDIN HAS A BLOCKED N-TERMINUS WITH MINOR SPECIES STARTING AT
CC       AMINO ACID 264 AND 273.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ---------------------------------------------------------------------
DR   EMBL; U58516; AAC50549.1; -.
DR   EMBL; S56151; AAB19771.1; -.
DR   MIM; 602281; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
FT   SIGNAL       1     23
FT   CHAIN       24    387       LACTADHERIN.
FT   CHAIN      202    387       LACTADHERIN; SHORT FORM
FT   CHAIN      268    317       MEDIN.
FT   DOMAIN      24     67       EGF LIKE.
FT   DOMAIN      70    225       F5/8 TYPE C 1.
FT   DOMAIN     230    387       F5/8 TYPE C 2.
FT   SITE        46     48       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    27     38       BY SIMILARITY.
FT   DISULFID    32     55       BY SIMILARITY.
FT   DISULFID    57     66       BY SIMILARITY.
FT   DISULFID    70    225       BY SIMILARITY.
FT   DISULFID   212    216       BY SIMILARITY.
FT   DISULFID   230    387       BY SIMILARITY.
FT   CARBOHYD   238    238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   325    325       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   329    329       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   350    350       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   387 AA;  43123 MW;  2EE857DEC8678B2D CRC64;


  Query Match             100.0%;  Score 53;  DB 1;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
            |||||||||
Db     271 KQGNFNAWV 279


RESULT   2
PSAB_LOTJA
ID   PSAB_LOTJA     STANDARD;      PRT;   734 AA.
AC   P58385;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Photosystem I P700 chlorophyll A apoprotein A2 (PsaB) (PSI-B).
GN   PSAB.
OS   Lotus japonicus.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Lotus.
OX   NCBI_TaxID=34305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN Accession MG-20;
RX   MEDLINE=21082929; PubMed=11214967;
RA   Kato T., Kaneko T., Sato S., Nakamura Y., Tabata S.;
RT   "Complete structure of the chloroplast genome of a legume, Lotus
RT   japonicus.";
RL   DNA Res. 7:323-330(2000).
CC   -!- FUNCTION: PsaA and psaB bind P700, the primary electron donor of
CC       photosystem I (PSI), as well as the electron acceptors A0, A1, and
CC       FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
CC       oxidoreductase.
CC   -!- COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
CC       is a phylloquinone and FX is a 4Fe-4S iron sulfur center.
CC   -!- SUBUNIT: A psaA/B heterodimer binds the P700 chlorophyll special
CC       pair and subsequent electron acceptors. The PSI reaction center of
CC       higher plants and algae is composed of one at least 11 subunits.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC       thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP002983; BAB33187.1; -.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Photosynthesis; Photosystem I; Electron transport; Chloroplast;
KW   Thylakoid; Transmembrane; Iron-sulfur; 4Fe-4S; Chlorophyll.
FT   TRANSMEM     46     69       I (POTENTIAL).
FT   TRANSMEM    135    158       II (POTENTIAL).
FT   TRANSMEM    175    199       III (POTENTIAL).
FT   TRANSMEM    273    291       IV (POTENTIAL).
FT   TRANSMEM    330    353       V (POTENTIAL).
FT   TRANSMEM    369    395       VI (POTENTIAL).
FT   TRANSMEM    417    439       VII (POTENTIAL).
FT   TRANSMEM    517    545       VIII (POTENTIAL).
FT   TRANSMEM    575    596       IX (POTENTIAL).
FT   TRANSMEM    643    665       X (POTENTIAL).
FT   TRANSMEM    707    727       XI (POTENTIAL).
FT   METAL       559    559       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   METAL       568    568       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   BINDING     654    654       P700 SPECIAL PAIR CHLOROPHYLL AXIAL
FT                                LIGAND (BY SIMILARITY).
FT   BINDING     662    662       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING     670    670       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING     671    671       A1 PHYLLOQUINONE (BY SIMILARITY).
FT   BINDING     700    700       A1 PHYLLOQUINONE (BY SIMILARITY).
SQ   SEQUENCE   734 AA;  82362 MW;  F83F2357E94D8DF6 CRC64;


  Query Match              79.2%;  Score 42;  DB 1;  Length 734;
  Best Local Similarity    87.5%;  Pred. No. 2.5;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
            |||| |||
Db      71 QGNFEAWV 78


RESULT   3
PSAB_OENHO
ID   PSAB_OENHO     STANDARD;      PRT;   734 AA.
```

```
AC   Q9MTN7;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Photosystem I P700 chlorophyll A apoprotein A2 (PsaB) (PSI-B).
GN   PSAB.
OS   Oenothera hookeri (Hooker's evening primrose).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots, Rosidae, eurosids II,
OC   Myrtales; Onagraceae; Oenothera.
OX   NCBI_TaxID=85636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Johansen;
RX   MEDLINE=20200218; PubMed=10952479;
RA   Hupfer H., Swiatek M., Hornung S., Herrmann R.G., Maier R.M.,
RA   Chiu W.-L., Sears B.;
RT   "Complete nucleotide sequence of the Oenothera elata plastid
RT   chromosome, representing plastome I of the five distinguishable
RT   Euoenothera plastomes.";
RL   Mol. Gen. Genet. 263:581-585(2000).
CC   -!- FUNCTION: PsaA and psaB bind P700, the primary electron donor of
CC       photosystem I (PSI), as well as the electron acceptors A0, A1, and
CC       FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
CC       oxidoreductase.
CC   -!- COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
CC       is a phylloquinone and FX is a 4Fe-4S iron-sulfur center.
CC   -!- SUBUNIT: A psaA/B heterodimer binds the P700 chlorophyll special
CC       pair and subsequent electron acceptors. The PSI reaction center of
CC       higher plants and algae is composed of one at least 11 subunits.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC       thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ271079; CAB67138.1; -.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Photosynthesis; Photosystem I; Electron transport; Chloroplast;
KW   Thylakoid; Transmembrane; Iron-sulfur; 4Fe-4S; Chlorophyll.
FT   TRANSMEM     46     69       I (POTENTIAL).
FT   TRANSMEM    135    157       II (POTENTIAL).
FT   TRANSMEM    175    199       III (POTENTIAL).
FT   TRANSMEM    273    291       IV (POTENTIAL).
FT   TRANSMEM    330    353       V (POTENTIAL).
FT   TRANSMEM    369    395       VI (POTENTIAL).
FT   TRANSMEM    417    439       VII (POTENTIAL).
FT   TRANSMEM    517    535       VIII (POTENTIAL).
FT   TRANSMEM    575    596       IX (POTENTIAL).
FT   TRANSMEM    643    665       X (POTENTIAL).
FT   TRANSMEM    707    727       XI (POTENTIAL).
FT   METAL       559    559       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   METAL       568    568       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   BINDING     654    654       P700 SPECIAL PAIR CHLOROPHYLL AXIAL
FT                                LIGAND (BY SIMILARITY).
FT   BINDING     662    662       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING     670    670       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING     671    671       A1 PHYLLOQUINONE (BY SIMILARITY).
FT   BINDING     700    700       A1 PHYLLOQUINONE (BY SIMILARITY).
SQ   SEQUENCE   734 AA;  82313 MW;  DBB713BBCC18ABDE CRC64;
```

```
  Query Match          79.2%;  Score 42;  DB 1;  Length 734;
  Best Local Similarity   87.5%;  Pred. No. 2.5;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QCNPNAWV 9
           |||| |||
Db      71 QGNPEAWV 78


RESULT   4
PSAB_PEA
ID   PSAB_PEA        STANDARD;       PRT:   734 AA.
AC   P05311;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Photosystem I P700 chlorophyll A apoprotein A2 (PsaB) (PSI-B).
GN   PSAB OR PSAA2.
OS   Pisum sativum (Garden pea).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae; Pisum.
OX   NCBI_TaxID=3888;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lehmbeck L., Rasmussen O.F., Bookjans G.B., Jepsen B.R.,
RA   Stummann B.M., Henningsen K.W.;
RT   "Sequence of two genes in pea chloroplast DNA coding for 84 and 82 kD
RT   polypeptides of the photosystem I complex.";
RL   Plant Mol. Biol. 7:33-40(1986).
CC   -!- FUNCTION: PsaA and psaB bind P700, the primary electron donor of
CC       photosystem I (PSI), as well as the electron acceptors A0, A1, and
CC       FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
CC       oxidoreductase.
CC   -!- COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
CC       is a phylloquinone and FX is a 4Fe-4S iron-sulfur center.
CC   -!- SUBUNIT: A psaA/B heterodimer binds the P700 chlorophyll special
CC       pair and subsequent electron acceptors. The PSI reaction center of
CC       higher plants and algae is composed of one at least 11 subunits.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC       thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X05423; CAA29004.1; -.
DR   PIR; S00704; S00704.
DR   Mendel; 5365; PISsa;psaB;1.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Photosynthesis; Photosystem I; Electron transport; Chloroplast;
KW   Thylakoid; Transmembrane; Iron-sulfur; 4Fe-4S; Chlorophyll.
FT   TRANSMEM     46     69       I (POTENTIAL).
FT   TRANSMEM    135    158       II (POTENTIAL).
FT   TRANSMEM    175    199       III (POTENTIAL).
FT   TRANSMEM    273    291       IV (POTENTIAL).
FT   TRANSMEM    330    353       V (POTENTIAL).
FT   TRANSMEM    369    395       VI (POTENTIAL).
FT   TRANSMEM    417    439       VII (POTENTIAL).
FT   TRANSMEM    517    535       VIII (POTENTIAL).
FT   TRANSMEM    575    596       IX (POTENTIAL).
FT   TRANSMEM    643    665       X (POTENTIAL).
FT   TRANSMEM    707    727       XI (POTENTIAL)
FT   METAL       559    559       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   METAL       568    568       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
```

```
FT   BINDING     654     654       P700 SPECIAL PAIR CHLOROPHYLL AXIAL
FT                                 LIGAND (BY SIMILARITY).
FT   BINDING     662     662       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING     670     670       A0 CHLOROPHYLL (BY SIMILARITY)
FT   BINDING     671     671       A1 PHYLLOQUINONE (BY SIMILARITY)
FT   BINDING     700     700       A1 PHYLLOQUINONE (BY SIMILARITY)
SQ   SEQUENCE   734 AA;  82412 MW;  2ACF1F77D46FA8E7 CRC64;


  Query Match             79.2%;  Score 42;  DB 1;  Length 734;
  Best Local Similarity   87.5%;  Pred. No. 2.5;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
            |||| |||
Db      71 QGNFEAWV 78


RESULT   5
PSAB_PINTH
ID   PSAB_PINTH     STANDARD;      PRT;   734 AA.
AC   P41640;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Photosystem I P700 chlorophyll A apoprotein A2 (PsaB) (PSI-B).
GN   PSAB.
OS   Pinus thunbergii (Green pine) (Japanese black pine).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX   NCBI_TaxID=3350;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95024047; PubMed=7937893;
RA   Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA   Sugiura M.;
RT   "Loss of all ndh genes as determined by sequencing the entire
RT   chloroplast genome of the black pine Pinus thunbergii.",
RL   Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC   -!- FUNCTION: PsaA and psaB bind P700, the primary electron donor of
CC       photosystem I (PSI), as well as the electron acceptors A0, A1, and
CC       FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
CC       oxidoreductase.
CC   -!- COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
CC       is a phylloquinone and FX is a 4Fe-4S iron-sulfur center.
CC   -!- SUBUNIT: A psaA/B heterodimer binds the P700 chlorophyll special
CC       pair and subsequent electron acceptors. The PSI reaction center of
CC       higher plants and algae is composed of one at least 11 subunits.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC       thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D17510; BAA04420.1; -.
DR   Mendel, 9995, PINth,psaB.1.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Photosynthesis, Photosystem I, Electron transport, Chloroplast,
KW   Thylakoid, Transmembrane, Iron-sulfur, 4Fe 4S, Chlorophyll.
FT   TRANSMEM     46     69       I (POTENTIAL).
FT   TRANSMEM    135    158       II (POTENTIAL).
FT   TRANSMEM    175    199       III (POTENTIAL).
FT   TRANSMEM    273    291       IV (POTENTIAL).
FT   TRANSMEM    330    353       V (POTENTIAL).
FT   TRANSMEM    369    395       VI (POTENTIAL).
FT   TRANSMEM    417    439       VII (POTENTIAL).
FT   TRANSMEM    517    535       VIII (POTENTIAL).
FT   TRANSMEM    575    596       IX (POTENTIAL).
FT   TRANSMEM    643    665       X (POTENTIAL).
FT   TRANSMEM    707    727       XI (POTENTIAL).
FT   METAL       559    559       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   METAL       568    568       IRON-SULFUR FX (4FE-4S).
FT   BINDING     654    654       P700 SPECIAL PAIR CHLOROPHYLL AXIAL
FT                                 LIGAND (BY SIMILARITY).
FT   BINDING     662    662       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING     670    670       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING     671    671       A1 PHYLLOQUINONE (BY SIMILARITY).
FT   BINDING     700    700       A1 PHYLLOQUINONE (BY SIMILARITY).
SQ   SEQUENCE   734 AA;  82495 MW;  083720BA7B19266F CRC64;


  Query Match             79.2%;  Score 42;  DB 1;  Length 734;
  Best Local Similarity   87.5%;  Pred. No. 2.5;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
            |||| |||
Db      71 QGNFEAWV 78


RESULT   6
PSAB_CYACA
ID   PSAB_CYACA     STANDARD;      PRT;   734 AA.
AC   Q9TLQ6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Photosystem I P700 chlorophyll A apoprotein A2 (PsaB) (PSI-B).
GN   PSAB.
OS   Cyanidium caldarium.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC   Cyanidium.
OX   NCBI_TaxID=2771;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RK-1;
RX   MEDLINE=20496459; PubMed=11040290;
RA   Gloeckner G., Rosenthal A., Valentin K.;
RT   "The structure and gene repertoire of an ancient red algal plastid
RT   genome.";
RL   J. Mol. Evol. 51:382-390(2000).
CC   -!- FUNCTION: PsaA and psaB bind P700, the primary electron donor of
CC       photosystem I (PSI), as well as the electron acceptors A0, A1, and
CC       FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
CC       oxidoreductase.
CC   -!- COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
CC       is a phylloquinone and FX is a 4Fe-4S iron-sulfur center.
CC   -!- SUBUNIT: A psaA/B heterodimer binds the P700 chlorophyll special
CC       pair and subsequent electron acceptors. The PSI reaction center of
CC       higher plants and algae is composed of one at least 11 subunits.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC       thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF022186; AAF12881.1; -.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
```

```
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Photosynthesis; Photosystem I; Electron transport; Chloroplast;
KW   Thylakoid; Transmembrane; Iron sulfur; 4Fe-4S; Chlorophyll.
FT   TRANSMEM      46     69       I (POTENTIAL).
FT   TRANSMEM     135    158       II (POTENTIAL).
FT   TRANSMEM     175    199       III (POTENTIAL).
FT   TRANSMEM     273    291       IV (POTENTIAL).
FT   TRANSMEM     330    353       V (POTENTIAL).
FT   TRANSMEM     369    395       VI (POTENTIAL).
FT   TRANSMEM     417    439       VII (POTENTIAL).
FT   TRANSMEM     517    535       VIII (POTENTIAL).
FT   TRANSMEM     575    596       IX (POTENTIAL).
FT   TRANSMEM     643    665       X (POTENTIAL).
FT   TRANSMEM     707    727       XI (POTENTIAL).
FT   METAL        559    559       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   METAL        568    568       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   BINDING      654    654       P700 SPECIAL PAIR CHLOROPHYLL AXIAL
FT                                 LIGAND (BY SIMILARITY).
FT   BINDING      662    662       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING      670    670       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING      671    671       A1 PHYLLOQUINONE (BY SIMILARITY).
FT   BINDING      700    700       A1 PHYLLOQUINONE (BY SIMILARITY).
SQ   SEQUENCE   734 AA;  82359 MW;  4496AA2AE59CA9B9 CRC64;


  Query Match             77.4%;  Score 41;  DB 1;  Length 734;
  Best Local Similarity   75.0%;  Pred. No. 3.8;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QCNPNAWV 9
           |||| ||:
Db      71 QGNFEAWI 78


RESULT   7
CYRG_MOUSE
ID   CYRG_MOUSE     STANDARD;      PRT:   369 AA.
AC   P34902;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Cytokine receptor common gamma chain precursor (Gamma-C)
DE   (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64).
GN   IL2RG.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93277575; PubMed=8504926;
RA   Kumaki S., Kondo M., Takeshita T., Asao H., Nakamura M., Sugamura K.;
RT   "Cloning of the mouse interleukin 2 receptor gamma chain:
RT   demonstration of functional differences between the mouse and human
RT   receptors.";
RL   Biochem. Biophys. Res. Commun. 193:356-363(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBA/CA;
RX   MEDLINE=93391374; PubMed=8378320;
RA   Cao X., Kozak C.A., Liu Y.J., Noguchi M., O'Connell E., Leonard W.J.;
RT   "Characterization of cDNAs encoding the murine interleukin 2 receptor
RT   (IL-2R) gamma chain: chromosomal mapping and tissue specificity of
RT   IL-2R gamma chain expression.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:8464-8468(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93366191; PubMed=8359699;
RA   Kobayashi N., Nakagawa S., Minami Y., Taniguchi T., Kono T.;
RT   "Cloning and sequencing of the cDNA encoding a mouse IL-2 receptor
RT   gamma.";
RL   Gene 130:303-304(1993).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95104285; PubMed=7805729;
RA   Disanto J.P., Cattuin S., Wilson A., Macdonald H.R., Avner P.,
RA   Fischer A., de Saint Basile G.;
RT   "The murine interleukin-2 receptor gamma chain gene: organization,
RT   chromosomal localization and expression in the adult thymus.";
RL   Eur. J. Immunol. 24:3014-3018(1994).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6.S;
RX   MEDLINE=96341745; PubMed=8750189;
RA   Chin R.K., Droll A., Cooper D.L., Dougherty S.T., Dirks J.F.,
RA   Dougherty G.J.;
RT   "Molecular mechanisms regulating the hyaluronan binding activity of
RT   the adhesion protein CD44.";
RL   J. Neuroimmunol. 66:231-239(1995).
CC   -!- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
CC       INTERLEUKINS.
CC   -!- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
CC       PROBABLY ALSO THE IL-13 RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D13821; BAA02974.1; -.
DR   EMBL; U21795; AAA64279.1; -.
DR   EMBL; D13565; BAA02760.1; -.
DR   EMBL; L20048; AAA39286.1; -.
DR   EMBL; S75852; AAB32904.1; -.
DR   EMBL; S75844; AAB32904.1; JOINED.
DR   EMBL; S75845; AAB32904.1; JOINED.
DR   EMBL; S75847; AAB32904.1; JOINED.
DR   EMBL; S75848; AAB32904.1; JOINED.
DR   EMBL; S75849; AAB32904.1; JOINED.
DR   EMBL; S75850; AAB32904.1; JOINED.
DR   EMBL; S75851; AAB32904.1; JOINED.
DR   EMBL; X75337; CAA53085.1; -.
DR   PIR; JN0592; JN0592.
DR   PIR; JN0775; JN0775.
DR   HSSP; P31785; 1ILN.
DR   MGD; MGI:96551; Il2rg.
DR   InterPro; IPR002996; CR1A.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003531; Hematopo_receptor_S_F1.
DR   Pfam; PF00041; fn3; 1.
DR   SMART; SM00060; FN3; 1.
DR   PROSITE; PS01355; HEMATOPO_REC_S_F1; 1.
KW   Receptor; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    369       CYTOKINE RECEPTOR COMMON GAMMA CHAIN.
FT   DOMAIN       23    263       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    264    284       POTENTIAL.
FT   DOMAIN      285    369       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      151    250       FIBRONECTIN TYPE-III.
FT   DISULFID     62     72       POTENTIAL.
FT   DISULFID    102    115       POTENTIAL.
FT   CARBOHYD     71     71       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     75     75       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     84     84       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     96     96       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    159    159       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    164    164       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   369 AA;  42241 MW;  CB2D5AB459077AC7 CRC64;
```

```
  Query Match             73.6%;  Score 39;  DB 1;  Length 369;
  Best Local Similarity   85.7%;  Pred. No. 4.5;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 QGNFNAW 8
           ||||:||
Db     304 QGNFSAW 310


RESULT   8
CYRG_BOVIN
ID   CYRG_BOVIN     STANDARD;      PRT;   379 AA.
AC   Q95118;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Cytokine receptor common gamma chain precursor (Gamma-C)
DE   (Interleukin-2 receptor gamma chain) (IL 2R gamma chain) (P64).
GN   IL2RG.
OS   Bos taurus (Bovine)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96268473; PubMed=8672241;
RA   Yoo J., Stone R.T., Solinas-Toldo S., Fries R., Beattie C.W.;
RT   "Cloning and chromosomal mapping of bovine interleukin-2 receptor
RT   gamma gene.";
RL   DNA Cell Biol. 15:453-459(1996).
CC   -!- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
CC       INTERLEUKINS
CC   -!- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
CC       PROBABLY ALSO THE IL-13 RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U33748; AAB07812.1; -.
DR   HSSP; P31785; 1ILN.
DR   InterPro; IPR002996; CR1A.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003531; Hematopo_receptor_S_F1.
DR   Pfam; PF00041; fn3; 1.
DR   SMART; SM00060; FN3; 1.
DR   PROSITE; PS01355; HEMATOPO_REC_S_F1; 1.
KW   Receptor; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23    379       CYTOKINE RECEPTOR COMMON GAMMA CHAIN.
FT   DOMAIN      23    269       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    270   290       POTENTIAL.
FT   DOMAIN      291   379       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      158   256       FIBRONECTIN TYPE-III.
FT   DISULFID     68    78       POTENTIAL.
FT   DISULFID    109   122       POTENTIAL.
FT   CARBOHYD     77    77       N-LINKED (GLCNAC . .) (POTENTIAL)
FT   CARBOHYD     81    81       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     90    90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    166   166       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    171   171       N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ   SEQUENCE   479 AA,  43037 MW,  83CFAD9C9B032178 CRC64;
```

```
  Query Match             73.6%;  Score 39;  DB 1;  Length 379;
  Best Local Similarity   85.7%;  Pred. No. 4.6;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 QGNFNAW 8
           ||||:||
Db     317 QGNFSAW 323


RESULT   9
PSAB_ANTMA
ID   PSAB_ANTMA     STANDARD;      PRT;   734 AA.
AC   Q33332;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Photosystem I P700 chlorophyll A apoprotein A2 (PsaB) (PSI-B).
GN   PSAB.
OS   Antirrhinum majus (Garden snapdragon).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Lamiales; Veronicaceae; Antirrhinum.
OX   NCBI_TaxID=4151;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SO; TISSUE=Leaf;
RX   MEDLINE=96133694; PubMed=8544819;
RA   Schaffner C., Laasch H., Hagemann R.;
RT   "Detection of point mutations in chloroplast genes of Antirrhinum
RT   majus L. I. Identification of a point mutation in the psaB gene of a
RT   photosystem I plastome mutant.";
RL   Mol. Gen. Genet. 249:533-544(1995).
CC   -!- FUNCTION: PsaA and psaB bind P700, the primary electron donor of
CC       photosystem I (PSI), as well as the electron acceptors A0, A1, and
CC       FX. PSI functions as a plastocyanin/cytochrome c6-terredoxin
CC       oxidoreductase.
CC   -!- COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
CC       is a phylloquinone and FX is a 4Fe-4S iron-sulfur center.
CC   -!- SUBUNIT: A psaA/B heterodimer binds the P700 chlorophyll special
CC       pair and subsequent electron acceptors. The PSI reaction center of
CC       higher plants and algae is composed of one at least 11 subunits.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC       thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X84153; CAA58958.1; -.
DR   Mendel; 13513, ANTma,psaB;2.
DR   InterPro; IPR001280, PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Photosynthesis, Photosystem I; Electron transport; Chloroplast;
KW   Thylakoid; Transmembrane; Iron-sulfur; 4Fe-4S; Chlorophyll.
FT   TRANSMEM     46     69       I (POTENTIAL).
FT   TRANSMEM    135    158       II (POTENTIAL).
FT   TRANSMEM    175    199       III (POTENTIAL).
FT   TRANSMEM    273    291       IV (POTENTIAL).
FT   TRANSMEM    330    353       V (POTENTIAL).
FT   TRANSMEM    369    395       VI (POTENTIAL).
FT   TRANSMEM    417    439       VII (POTENTIAL).
FT   TRANSMEM    517    535       VIII (POTENTIAL).
FT   TRANSMEM    575    596       IX (POTENTIAL).
FT   TRANSMEM    643    665       X (POTENTIAL).
FT   TRANSMEM    707    727       XI (POTENTIAL).
```

```
FT   METAL       559     559       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   METAL       568     568       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   BINDING     654     654       P700 SPECIAL PAIR CHLOROPHYLL AXIAL
FT                                 LIGAND (BY SIMILARITY).
FT   BINDING     662     662       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING     670     670       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING     671     671       A1 PHYLLOQUINONE (BY SIMILARITY).
FT   BINDING     700     700       A1 PHYLLOQUINONE (BY SIMILARITY).
SQ   SEQUENCE   734 AA;  82429 MW;  AE9DB0B0DF235407 CRC64;


  Query Match             73.6%;  Score 39;  DB 1;  Length 734;
  Best Local Similarity   75.0%;  Pred. No. 9;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
           |||| :||
Db      71 QGNFESWV 78


RESULT  10
PSAB_SPIOL
ID   PSAB_SPIOL     STANDARD;      PRT;   734 AA.
AC   P06512;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Photosystem I P700 chlorophyll A apoprotein A2 (PsaB) (PSI-B).
GN   PSAB.
OS   Spinacia oleracea (Spinach).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX   NCBI_TaxID=3562;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kirsch W., Seyer P., Herrmann R.G.;
RT   "Nucleotide sequence of the clustered genes for two P700 chlorophyll
RT   a apoproteins of the photosystem I reaction center and the ribosomal
RT   protein S14 of the spinach plastid chromosome.";
RL   Curr. Genet. 10:843-855(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. CHANT D'HIVER, AND CV. MONATOL;
RX   MEDLINE=21187424; PubMed=11292076;
RA   Schmitz-Linneweber C., Maier R.M., Alcaraz J.-P., Cottet A.,
RA   Herrmann R.G., Mache R.;
RT   "The plastid chromosome of spinach (Spinacia oleracea): complete
RT   nucleotide sequence and gene organization.";
RL   Plant Mol. Biol. 45:307-315(2001).
CC   -!- FUNCTION: PsaA and psaB bind P700, the primary electron donor of
CC       photosystem I (PSI), as well as the electron acceptors A0, A1, and
CC       FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
CC       oxidoreductase.
CC   -!- COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
CC       is a phylloquinone and FX is a 4Fe-4S iron-sulfur center.
CC   -!- SUBUNIT: A psaA/B heterodimer binds the P700 chlorophyll special
CC       pair and subsequent electron acceptors. The PSI reaction center of
CC       higher plants and algae is composed of one at least 11 subunits.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC       thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ400848; CAB88725.1.  -
DR   PIR; S00445; S00445.
DR   Mendel; 2573; SPIol;psaB;1.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Photosynthesis; Photosystem I; Electron transport; Chloroplast;
KW   Thylakoid; Transmembrane; Iron-sulfur; 4Fe-4S; Chlorophyll.
FT   TRANSMEM     46      69       I (POTENTIAL).
FT   TRANSMEM    135     158       II (POTENTIAL).
FT   TRANSMEM    175     199       III (POTENTIAL).
FT   TRANSMEM    273     291       IV (POTENTIAL).
FT   TRANSMEM    330     353       V (POTENTIAL).
FT   TRANSMEM    369     395       VI (POTENTIAL).
FT   TRANSMEM    417     439       VII (POTENTIAL).
FT   TRANSMEM    517     535       VIII (POTENTIAL).
FT   TRANSMEM    575     596       IX (POTENTIAL).
FT   TRANSMEM    643     665       X (POTENTIAL).
FT   TRANSMEM    707     727       XI (POTENTIAL).
FT   METAL       559     559       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   METAL       568     568       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   BINDING     654     654       P700 SPECIAL PAIR CHLOROPHYLL AXIAL
FT                                 LIGAND (BY SIMILARITY).
FT   BINDING     662     662       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING     670     670       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING     671     671       A1 PHYLLOQUINONE (BY SIMILARITY).
FT   BINDING     700     700       A1 PHYLLOQUINONE (BY SIMILARITY).
SQ   SEQUENCE   734 AA;  82429 MW;  9C4B83803338C12F CRC64;


  Query Match             73.6%;  Score 39;  DB 1;  Length 734;
  Best Local Similarity   75.0%;  Pred. No. 9;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
           |||| :||
Db      71 QGNFESWV 78


RESULT  11
PSAB_TOBAC
ID   PSAB_TOBAC     STANDARD;      PRT;   734 AA.
AC   P06407;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Photosystem I P700 chlorophyll A apoprotein A2 (PsaB) (PSI-B).
GN   PSAB.
OS   Nicotiana tabacum (Common tobacco).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids 1; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. BRIGHT YELLOW 4;
RA   Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N.,
RA   Matsubayashi T., Zaita N., Chunwongse J., Obokata J.,
RA   Yamaguchi-Shinozaki K., Ohto C., Torazawa K., Meng B.Y., Sugita M.,
RA   Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A.,
RA   Tohdoh N., Shimada H., Sugiura M.;
RT   "The complete nucleotide sequence of the tobacco chloroplast genome:
RT   its gene organization and expression.";
RL   EMBO J. 5:2043-2049(1986).
CC   -!- FUNCTION: PsaA and psaB bind P700, the primary electron donor of
CC       photosystem I (PSI), as well as the electron acceptors A0, A1, and
CC       FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
CC       oxidoreductase.
CC   -!- COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
CC       is a phylloquinone and FX is a 4Fe-4S iron-sulfur center.
CC   -!- SUBUNIT: A psaA/B heterodimer binds the P700 chlorophyll special
CC       pair and subsequent electron acceptors. The PSI reaction center of
```

```
CC       higher plants and algae is composed of one at least 11 subunits.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC       thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z00044; CAA77351.1; -.
DR   PIR; A03466; A2NTP7.
DR   Mendel; 9442; Nicta psaB 1
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Photosynthesis; Photosystem I; Electron transport; Chloroplast;
KW   Thylakoid; Transmembrane; Iron-sulfur; 4Fe-4S; Chlorophyll.
FT   TRANSMEM      46     69       I (POTENTIAL).
FT   TRANSMEM     135    158       II (POTENTIAL).
FT   TRANSMEM     175    199       III (POTENTIAL).
FT   TRANSMEM     273    291       IV (POTENTIAL).
FT   TRANSMEM     330    353       V (POTENTIAL).
FT   TRANSMEM     369    395       VI (POTENTIAL).
FT   TRANSMEM     417    439       VII (POTENTIAL).
FT   TRANSMEM     517    535       VIII (POTENTIAL).
FT   TRANSMEM     575    596       IX (POTENTIAL).
FT   TRANSMEM     643    665       X (POTENTIAL).
FT   TRANSMEM     707    727       XI (POTENTIAL).
FT   METAL        559    559       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   METAL        568    568       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   BINDING      654    654       P700 SPECIAL PAIR CHLOROPHYLL AXIAL
FT                                 LIGAND (BY SIMILARITY).
FT   BINDING      662    662       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING      670    670       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING      671    671       A1 PHYLLOQUINONE (BY SIMILARITY).
FT   BINDING      700    700       A1 PHYLLOQUINONE (BY SIMILARITY).
SQ   SEQUENCE   734 AA;  82409 MW;  0244DAB0E5414BFF CRC64;
```

```
  Query Match             73.6%;  Score 39;  DB 1;  Length 734;
  Best Local Similarity   75.0%;  Pred. No. 9,
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
           |||| :||
Db      71 QGNFESWV 78
```

```
RESULT 12
MFGM_PIG
ID   MFGM_PIG        STANDARD;      PRT;   409 AA.
AC   P79385;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Lactadherin (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM) (Sperm
DE   surface protein SP47) (PP47).
GN   MFGE8.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
```

```
CC       BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC       SPERMATOZOAN.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y11684; CAA72379.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Glycoprotein; Repeat; EGF-like domain.
FT   DOMAIN        2     41       EGF-LIKE 1.
FT   DOMAIN       44     88       EGF-LIKE 2.
FT   DOMAIN       91    247       F5/8 TYPE C 1.
FT   DOMAIN      252    409       F5/8 TYPE C 2.
FT   SITE         67     69       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID      6     17       BY SIMILARITY.
FT   DISULFID     11     29       BY SIMILARITY.
FT   DISULFID     31     40       BY SIMILARITY.
FT   DISULFID     91    247       BY SIMILARITY.
FT   DISULFID    234    238       BY SIMILARITY.
FT   DISULFID    252    409       BY SIMILARITY.
FT   CARBOHYD     41     41       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    372    372       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   409 AA;  45726 MW;  B0B0B7AFB0B29927A CRC64;
```

```
  Query Match             71.7%;  Score 38;  DB 1;  Length 409;
  Best Local Similarity   85.7%;  Pred. No. 7.7;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAW 8
           || ||||
Db     294 QGKFNAW 300
```

```
RESULT 13
MFGM_BOVIN
ID   MFGM_BOVIN      STANDARD;      PRT;   427 AA.
AC   Q95114; Q27959; P79344;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
DE   (MGP57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
DE   SP47) (BP47) (Components 15/16).
GN   MFGE8.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC   STRAIN=HOLSTEIN; TISSUE=Mammary gland;
RX   MEDLINE=97008954; PubMed=8856064;
```

```
PA   Hvarregaard I., Andersen M.H., Berglund L., Rasmussen J.T.,
RA   Petersen T.E.;
RT   "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT   milk fat globules.";
RL   Eur. J. Biochem. 240:628-636(1996).
RN   [2]
PP   SEQUENCE OF 18-427 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=96125736; PubMed=8541316;
RA   Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA   Matsuda T.;
RT   "Molecular cloning of glycoprotein antigens MGF57/53 recognized by
RT   monoclonal antibodies raised against bovine milk fat globule
RT   membrane.";
RL   Biochim. Biophys. Acta 1245:385-391(1995).
RN   [3]
PP   SEQUENCE OF 19-427 FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
PP   SEQUENCE OF 140-146; 174-187; 233-245 AND 422-427.
RC   TISSUE=Milk;
RX   MEDLINE=93250576; PubMed=8485470;
RA   Mather I.H., Banghart L.R., Lane W.S.;
RT   "The major fat-globule membrane proteins, bovine components 15/16 and
RT   guinea-pig GP 55, are homologous to MGF-E8, a murine glycoprotein
RT   containing epidermal growth factor-like and factor V/VIII-like
RT   sequences.";
RL   Biochem. Mol. Biol. Int. 29:545-554(1993).
CC   -!- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
CC       MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM
CC       LACKS 53 AMINO ACIDS WITHIN THE F5/8 TYPE C 1 DOMAIN
CC   -!- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN
CC   -!- PTM: THE 2 O-LINKED GLYCANS CONSTIST OF GAL, GLCNAC AND FUC, WITH
CC       PROBLABLY FUC AS REDUCING TERMINAL SUGAR.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------- ---------  ----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X91895; CAA62997.1; -.
DR   EMBL; S80643; AAB35894.2; -.
DR   EMBL; Y11719; CAA72406.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain;
KW   Alternative splicing.
FT   SIGNAL       1     18
FT   CHAIN       19    427       LACTADHERIN.
FT   DOMAIN      20     59       EGF-LIKE 1.
FT   DOMAIN      62    106       EGF-LIKE 2.
FT   DOMAIN     109    265       F5/8 TYPE C 1
FT   DOMAIN     270    427       F5/8 TYPE C 2
FT   SITE        85     87       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    24     45       BY SIMILARITY.
FT   DISULFID    29     47       BY SIMILARITY.
FT   DISULFID    49     58       BY SIMILARITY.
FT   DISULFID    66     77       BY SIMILARITY.
FT   DISULFID    71     94       BY SIMILARITY.
FT   DISULFID    96    105       BY SIMILARITY.
FT   DISULFID   109    265
FT   DISULFID   252    256
FT   DISULFID   270    427
FT   CARBOHYD    27     27       O-LINKED (FUC. . .) (IN PAS-6).
FT   CARBOHYD    34     34       O-LINKED (FUC. . .) (IN PAS-7).
FT   CARBOHYD    59     59       N-LINKED (GLCNAC. . .) (HYBRID; (IN PAS-6
FT                               AND PAS-7).
FT   CARBOHYD   227    227       N-LINKED (GLCNAC. . .) (HIGH MANNOSE) (IN
FT                               PAS-6).
FT   VARSPLIC   169    221       MISSING (IN SHORT ISOFORM).
FT   CONFLICT    19     19       A -> F (IN REF. 1).
FT   CONFLICT    28     28       L -> Q (IN REF. 1).
SQ   SEQUENCE   427 AA;  47411 MW;  4CBBEE3A1DC4EB24 CRC64;
```

```
  Query Match             71.7%;  Score 38;  DB 1;  Length 427;
  Best Local Similarity   85.7%;  Pred. No. 8;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNPNAW 8
           || ||||
Db     312 QGKPNAW 318
```

```
RESULT 14
PSAB_ARATH
ID   PSAB_ARATH     STANDARD;      PRT;   734 AA.
AC   P56767;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Photosystem I P700 chlorophyll A apoprotein A2 (PsaB) (PSI-B).
GN   PSAB.
OS   Arabidopsis thaliana (Mouse-ear cress)
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. COLUMBIA;
RX   MEDLINE=20039611; PubMed=10574454;
RA   Sato S., Nakamura Y., Kaneko T., Asamizu E., Tabata S.;
RT   "Complete structure of the chloroplast genome of Arabidopsis
RT   thaliana.";
RL   DNA Res. 6:283-290(1999).
CC   -!- FUNCTION: PsaA and psaB bind P700, the primary electron donor of
CC       photosystem I (PSI), as well as the electron acceptors A0, A1, and
CC       FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
CC       oxidoreductase.
CC    !  COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
CC       is a phylloquinone and FX is a 4Fe-4S iron sulfur center.
CC    !  SUBUNIT: A psaA/B heterodimer binds the P700 chlorophyll special
CC       pair and subsequent electron acceptors. The PSI reaction center of
CC       higher plants and algae is composed of one at least 11 subunits.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC       thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; AP000423; BAA84384.1; -.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Photosynthesis; Photosystem I; Electron transport; Chloroplast;
KW   Thylakoid, Transmembrane, Iron sulfur, 4Fe 4S, Chlorophyll
FT   TRANSMEM      46     69       I (POTENTIAL).
FT   TRANSMEM     135    158       II (POTENTIAL).
FT   TRANSMEM     175    199       III (POTENTIAL).
FT   TRANSMEM     273    291       IV (POTENTIAL).
FT   TRANSMEM     330    353       V (POTENTIAL).
FT   TRANSMEM     369    395       VI (POTENTIAL).
FT   TRANSMEM     417    439       VII (POTENTIAL).
FT   TRANSMEM     517    535       VIII (POTENTIAL).
FT   TRANSMEM     575    596       IX (POTENTIAL).
FT   TRANSMEM     643    665       X (POTENTIAL).
FT   TRANSMEM     707    727       XI (POTENTIAL).
FT   METAL        559    559       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   METAL        568    568       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   BINDING      654    654       P700 SPECIAL PAIR CHLOROPHYLL AXIAL
FT                                 LIGAND (BY SIMILARITY).
FT   BINDING      662    662       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING      670    670       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING      671    671       A1 PHYLLOQUINONE (BY SIMILARITY).
FT   BINDING      700    700       A1 PHYLLOQUINONE (BY SIMILARITY).
SQ   SEQUENCE   734 AA;  82475 MW;  8D6A22BCBC6E37FC CRC64;


  Query Match             71.7%;  Score 38;  DB 1;  Length 734;
  Best Local Similarity   75.0%;  Pred. No. 14;
  Matches    6;  Conservative    0,  Mismatches    2,  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
            ||||  ||
Db      71 QGNFETWV 78


RESULT  15
PSAB_MARPO
ID   PSAB_MARPO     STANDARD;      PRT;   734 AA.
AC   P06408;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Photosystem I P700 chlorophyll A apoprotein A2 (PsaB) (PSI-B)
GN   PSAB.
OS   Marchantia polymorpha (Liverwort).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC   Marchantiales; Marchantiaceae; Marchantia.
OX   NCBI_TaxID=3197;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA   Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA   Ozeki H.;
RT   "Chloroplast gene organization deduced from complete sequence of
RT   liverwort Marchantia polymorpha chloroplast DNA.";
RL   Nature 322:572-574(1986).
RN   [2]
RP   SEQUENCE OF 591-734 FROM N.A.
RX   MEDLINE=85087956; PubMed=6393057;
RA   Umesono K., Inokuchi H., Ohyama K., Ozeki H.;
RT   "Nucleotide sequence of Marchantia polymorpha chloroplast DNA: a
RT   region possibly encoding three tRNAs and three proteins including a
RT   homologue of E. coli ribosomal protein S14.";
RL   Nucleic Acids Res. 12:9551-9565(1984).
CC   -!- FUNCTION: PsaA and psaB bind P700, the primary electron donor of
CC       photosystem I (PSI), as well as the electron acceptors A0, A1, and
CC       FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
CC       oxidoreductase.
CC   -!- COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
CC       is a phylloquinone and FX is a 4Fe-4S iron-sulfur center.
CC   -!- SUBUNIT: A psaA/B heterodimer binds the P700 chlorophyll special
CC       pair and subsequent electron acceptors. The PSI reaction center of
CC       higher plants and algae is composed of one at least 11 subunits.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC       thylakoid membrane.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X04465; CAA28084.1; -.
DR   EMBL; X01647; CAA25804.1; -.
DR   PIR; A03467; A2LVP7.
DR   PIR; S01605; S01605.
DR   Mendel; 2570; MARpo;psaB;1.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Photosynthesis; Photosystem I; Electron transport; Chloroplast;
KW   Thylakoid, Transmembrane, Iron-sulfur, 4Fe 4S, Chlorophyll.
FT   TRANSMEM      46     69       I (POTENTIAL).
FT   TRANSMEM     135    158       II (POTENTIAL).
FT   TRANSMEM     175    199       III (POTENTIAL).
FT   TRANSMEM     273    291       IV (POTENTIAL).
FT   TRANSMEM     330    353       V (POTENTIAL).
FT   TRANSMEM     369    395       VI (POTENTIAL).
FT   TRANSMEM     417    439       VII (POTENTIAL).
FT   TRANSMEM     517    535       VIII (POTENTIAL).
FT   TRANSMEM     575    596       IX (POTENTIAL).
FT   TRANSMEM     643    665       X (POTENTIAL).
FT   TRANSMEM     707    727       XI (POTENTIAL).
FT   METAL        559    559       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   METAL        568    568       IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
FT   BINDING      654    654       P700 SPECIAL PAIR CHLOROPHYLL AXIAL
FT                                 LIGAND (BY SIMILARITY).
FT   BINDING      662    662       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING      670    670       A0 CHLOROPHYLL (BY SIMILARITY).
FT   BINDING      671    671       A1 PHYLLOQUINONE (BY SIMILARITY).
FT   BINDING      700    700       A1 PHYLLOQUINONE (BY SIMILARITY).
SQ   SEQUENCE   734 AA;  82415 MW;  6E009560ECCCD986 CRC64;


  Query Match             71.7%;  Score 38;  DB 1;  Length 734;
  Best Local Similarity   85.7%;  Pred. No. 14;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAW 8
            |||| ||
Db      71 QGNFEAW 77


Search completed: September  5, 2002, 15:41:39
Job time: 474 sec
```

ConCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:23:15 ; Search time 122.86 Seconds
(without alignments)
12.673 Million cell updates/sec

Title: US-09-744-804-35
Perfect score: 53
Sequence: 1 KQGNFNAWV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172961929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	53	100.0	335	4	Q9BTL9	Q9btl9 homo sapien
2	43	81.1	363	6	O77718	O77718 equus caball
3	42	79.2	282	8	Q9TNK3	Q9tnk3 anthoceros
4	42	79.2	591	8	Q9TNI6	Q9tni9 sphagnum ta
5	42	79.2	653	2	Q9PDB5	Q9pdb5 streptococ
6	39	73.6	135	8	Q33331	Q33331 antirrhinum
7	38	71.7	64	4	P77085	P77085 escherichia
8	38	71.7	291	8	Q9TNI4	Q9tni4 physcomitre
9	38	71.7	301	8	Q9TNK9	Q9tnk9 coleochaete
10	38	71.7	586	8	Q9TNI9	Q9tni9 adiantum ca
11	38	71.7	604	8	Q9TNJ4	Q9tnj4 haplomitriu
12	37	69.8	297	10	Q94BV6	Q94bv6 arabidopsis
13	37	69.8	345	10	Q9M3U4	Q9m3u4 vitis vinif
14	37	69.8	379	10	Q941C7	Q941c7 arabidopsis
15	37	69.8	502	3	Q07429	Q07429 saccharomyc
16	37	69.8	817	16	P73619	P73619 synechocyst
17	37	69.8	2432	12	Q91QP4	Q91qp4 aichi virus
18	37	69.8	2433	12	Q91464	Q91464 aichi virus
19	36	67.9	126	11	Q9QZJ6	Q9qzj6 mus musculu
20	36	67.9	135	15	Q9EP38	Q9ep38 equine infe
21	36	67.9	135	15	Q99E51	Q99e51 equine infe
22	36	67.9	135	15	Q99EL9	Q99el9 equine infe
23	36	67.9	135	15	Q99DX1	Q99dx1 equine infe
24	36	67.9	257	2	Q9RBU1	Q9rbu1 alpha prote
25	36	67.9	295	5	Q9U9A3	Q9u9a3 dictyosteli
26	36	67.9	345	2	P94774	P94774 erwinia chr
27	36	67.9	359	11	Q9JJH8	Q9jjh8 mus musculu
28	36	67.9	359	11	Q9JMD5	Q9jmd5 mus musculu
29	36	67.9	359	11	Q9JIE7	Q9jie7 mus musculu
30	36	67.9	470	11	Q9JIQ7	Q9jiq7 mus musculu
31	36	67.9	471	4	Q9HC73	Q9hc73 homo sapien
32	36	67.9	480	4	O43854	O43854 homo sapien
33	36	67.9	480	11	O35474	O35474 mus musculu
34	36	67.9	482	5	Q9XY78	Q9xy78 acanthamoeb
35	36	67.9	526	16	Q9RDT8	Q9rdt8 bacillus ha
36	36	67.9	928	11	Q99N72	Q99n72 mus musculu
37	36	67.9	1564	4	O15077	O15077 homo sapien
38	35	66.0	81	15	Q9IV89	Q9iv89 human immun
39	35	66.0	135	15	Q99E91	Q99e91 equine infe
40	35	66.0	147	2	Q57080	Q57080 salmonella
41	35	66.0	147	2	Q57351	Q57351 salmonella
42	35	66.0	147	2	Q54055	Q54055 salmonella
43	35	66.0	147	2	Q54043	Q54043 salmonella
44	35	66.0	147	2	Q54045	Q54045 salmonella
45	35	66.0	147	2	Q54047	Q54047 salmonella

ALIGNMENTS

RESULT 1
Q9BTL9
ID Q9BTL9 PRELIMINARY; PRT; 335 AA.
AC Q9BTL9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO MILK FAT GLOBULE EGF FACTOR 8 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA.;
RA Strausberg R.;
RL Submitted (FEB 2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
DR EMBL; BC003610; AAH03610.1; -.
DR HSSP; P08709; 1BF9.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000421; FA58_C.
DR PRINTS; PR00010; EGFBLOOD.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00135; EGF_1; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01285; FA58C_1; 1.
DR PROSITE; PS01286; FA58C_2; 2.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 335 AA; 37523 MW; 20E84E5DB78E8E25 CRC64;

Query Match 100.0%; Score 53; DB 4; Length 335;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 KQGNFNAWV 9
             |||||||||
Db     271 KQGNFNAWV 279


RESULT   2
O77718
ID   O77718      PRELIMINARY;      PRT:   363 AA.
AC   O77718;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SPERM-MEMBRANE ASSOCIATED PROTEIN P47 (FRAGMENT).
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Gentzel M., Toepfer-Petersen E.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ010121; CAA09010.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 1.
KW   EGF like domain; Glycoprotein.
FT   NON_TER       1       1
FT   NON_TER     363     363
SQ   SEQUENCE   363 AA;  40744 MW;  1F8B6395AF32338D CRC64;


  Query Match             81.1%;  Score 43;  DB 6;  Length 363;
  Best Local Similarity   87.5%;  Pred. No. 6.5;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAW 8
             ||| ||||
Db     254 KQGKFNAW 261


RESULT   3
Q9TNK3
ID   Q9TNK3      PRELIMINARY;      PRT:   282 AA.
AC   Q9TNK3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PHOTOSYSTEM I P700 APOPROTEIN A2 (FRAGMENT).
GN   PSAB.
OS   Anthoceros punctatus.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Anthocerotophyta;
OC   Anthocerotales; Anthocerotaceae; Anthoceros.
OX   NCBI_TaxID=3234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CALLUS;
RX   MEDLINE=99404147; PubMed=10474899;
RA   Nishiyama T., Kato M.;
RT   "Molecular phylogenetic analysis among bryophytes and tracheophytes
RT   based on combined data of plastid coded genes and the 18S rRNA gene.";
RL   Mol. Biol. Evol. 16:1027-1036(1999).
DR   EMBL; AB013665; BAA83441.1; -.
DR   HSSP; P25896; 1JB0.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
KW   Chloroplast.
FT   NON_TER     282     282
SQ   SEQUENCE   282 AA;  31691 MW;  D4C36B58508B6044 CRC64;


  Query Match             79.2%;  Score 42;  DB 8;  Length 282;
  Best Local Similarity   87.5%;  Pred. No. 7.5;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
             |||| |||
Db      71 QGNFEAWV 78


RESULT   4
Q9TNJ9
ID   Q9TNJ9      PRELIMINARY;      PRT:   591 AA.
AC   Q9TNJ9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PHOTOSYSTEM I P700 APOPROTEIN A2 (FRAGMENT).
GN   PSAB.
OS   Sphagnum fallax.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC   Sphagnopsida; Sphagnales; Sphagnaceae; Sphagnum.
OX   NCBI_TaxID=53036;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99404147; PubMed=10474899;
RA   Nishiyama T., Kato M.;
RT   "Molecular phylogenetic analysis among bryophytes and tracheophytes
RT   based on combined data of plastid coded genes and the 18S rRNA gene.";
RL   Mol. Biol. Evol. 16:1027-1036(1999).
DR   EMBL; AB013671; BAA83447.1; -.
DR   HSSP; P25896; 1JB0.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Chloroplast.
FT   NON_TER     591     591
SQ   SEQUENCE   591 AA;  65591 MW;  3715F744A52D1442 CRC64;


  Query Match             79.2%;  Score 42;  DB 8;  Length 591;
  Best Local Similarity   87.5%;  Pred. No. 17;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
             |||| |||
Db      71 QGNFEAWV 78


RESULT   5
Q9RDB5
ID   Q9RDB5      PRELIMINARY;      PRT:   653 AA.
AC   Q9RDB5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE GLYCOSYL TRANSFERASE.
GN   SCE20.10.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
```

```
RA   Seeger K.J., Harris D.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL136058; CAB65566.1; -.
DR   InterPro; IPR001173; Glycos_transf_2.
DR   Pfam; PF00535; Glycos_transf_2; 1.
KW   Transferase.
SQ   SEQUENCE   653 AA;  74160 MW;  97F92D8AB06B1A4C CRC64;


  Query Match             79.2%;  Score 42;  DB 2;  Length 653;
  Best Local Similarity   66.7%;  Pred. No. 19;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KQGNFNAWV 9
            | ||:|||:
Db     239 KHGNYNAWL 247


RESULT   6
Q33331
ID   Q33331      PRELIMINARY;      PRT;   135 AA.
AC   Q33331;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PHOTOSYSTEM I SUBUNIT.
GN   PSA B.
OS   Antirrhinum majus (Garden snapdragon).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Lamiales; Veronicaceae; Antirrhinum.
OX   NCBI_TaxID=4151;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEAF;
RX   MEDLINE=96133694; PubMed=8544819;
RA   Schaffner C., Laasch H., Hagemann R.;
RT   "Detection of point mutations in chloroplast genes of Antirrhinum
RT   majus L. 1.Identification of a point mutation in the psaB gene of a
RT   photosystem I plastome mutant ";
RL   Mol. Gen. Genet. 249:533-544(1995).
DR   EMBL; X84152; CAA58957.1; -.
DR   HSSP; P25896; 1JB0.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
KW   Chloroplast.
SQ   SEQUENCE   135 AA;  15530 MW;  9CDBF9DFD6CADBF2 CRC64;


  Query Match             73.6%;  Score 39;  DB 8;  Length 135;
  Best Local Similarity   75.0%;  Pred. No. 12;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
            |||| :||
Db      71 QGNFESWV 78


RESULT   7
P77085
ID   P77085      PRELIMINARY;      PRT;   64 AA.
AC   P77085;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL PROTEIN.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA   Federspiel N., Hyman R., Kalman S., Kump C., Kurdi O., Lew H., Lin D.,
RA   Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RT   "Sequence of minutes 4-25 of Escherichia coli.";
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U82598; AAB40755.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   64 AA;  7520 MW;  7225B9ED5E6B92B7 CRC64;


  Query Match             71.7%;  Score 38;  DB 2;  Length 64;
  Best Local Similarity   87.5%;  Pred. No. 8.1;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 GNFNAWV 9
            |||:|||
Db      35 GNFSAWV 41


RESULT   8
Q9TNL4
ID   Q9TNL4      PRELIMINARY;      PRT;   291 AA.
AC   Q9TNL4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PHOTOSYSTEM I P700 APOPROTEIN A2 (FRAGMENT).
GN   PSAB.
OS   Physcomitrella patens (Moss).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC   Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
OX   NCBI_TaxID=3218;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99404147; PubMed=10474899;
RA   Nishiyama T., Kato M.;
RT   "Molecular phylogenetic analysis among bryophytes and tracheophytes
RT   based on combined data of plastid coded genes and the 18S rRNA gene.";
RL   Mol. Biol. Evol. 16:1027-1036(1999).
DR   EMBL; AB013653; BAA83429.1; -.
DR   HSSP; P25896; 1JB0.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
KW   Chloroplast.
FT   NON_TER     291    291
SQ   SEQUENCE   291 AA;  32606 MW;  8FB9F7179F1E01DB CRC64;


  Query Match             71.7%;  Score 38;  DB 8;  Length 291;
  Best Local Similarity   85.7%;  Pred. No. 41;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAW 8
            |||| ||
Db      71 QGNFEAW 77
```

```
RESULT   9
Q9TNK9
ID   Q9TNK9      PRELIMINARY;      PRT;   301 AA.
AC   Q9TNK9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PHOTOSYSTEM I P700 APOPROTEIN A2 (FRAGMENT).
GN   PSAB.
OS   Coleochaete nitellarum.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Coleochaetaceae; Coleochaete.
OX   NCBI_TaxID=78178;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN UTEX LB1261;
RX   MEDLINE=99404147; PubMed=10474899;
RA   Nishiyama T., Kato M.;
RT   "Molecular phylogenetic analysis among bryophytes and tracheophytes
RT   based on combined data of plastid coded genes and the 18S rRNA gene.";
RL   Mol. Biol. Evol. 16:1027-1036(1999)
DR   EMBL; AB013659; BAA83435.1; -.
DR   HSSP; P25896; 1JB0.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
KW   Chloroplast.
FT   NON_TER     301     301
SQ   SEQUENCE   301 AA;  33636 MW;  2572D2DA7D09247E CRC64;


  Query Match             71.7%;  Score 38;  DB 8;  Length 301;
  Best Local Similarity   85.7%;  Pred. No. 43;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAW 8
           |||| ||
Db      71 QGNFEAW 77


RESULT  10
Q9TNI9
ID   Q9TNI9      PRELIMINARY;      PRT;   586 AA.
AC   Q9TNI9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PHOTOSYSTEM I P700 APOPROTEIN A2 (FRAGMENT).
GN   PSAB.
OS   Adiantum capillus-veneris (Fern).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Filicophyta; Filicopsida; Filicales; Adiantaceae; Adiantum.
OX   NCBI_TaxID=13818;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99404147; PubMed=10474899;
RA   Nishiyama T., Kato M.;
RT   "Molecular phylogenetic analysis among bryophytes and tracheophytes
RT   based on combined data of plastid coded genes and the 18S rRNA gene.";
RL   Mol. Biol. Evol. 16:1027-1036(1999)
DR   EMBL; AB013681; BAA83457.1; -.
DR   HSSP; P25896; 1JB0.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Chloroplast.
FT   NON_TER     586     586
SQ   SEQUENCE   586 AA;  64958 MW;  8A03C5B08BFCFB68 CRC64;


  Query Match             71.7%;  Score 38;  DB 8;  Length 586;
  Best Local Similarity   75.0%;  Pred. No. 88;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 QGNFNAWV 9
           ||||  ||
Db      71 QGNFETWV 78


RESULT  11
Q9TNJ4
ID   Q9TNJ4      PRELIMINARY;      PRT;   604 AA.
AC   Q9TNJ4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PHOTOSYSTEM I P700 APOPROTEIN A2 (FRAGMENT).
GN   PSAB.
OS   Haplomitrium mnioides.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC   Calobryales; Haplomitriaceae; Haplomitrium.
OX   NCBI_TaxID=56921;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99404147; PubMed=10474899;
RA   Nishiyama T., Kato M.;
RT   "Molecular phylogenetic analysis among bryophytes and tracheophytes
RT   based on combined data of plastid coded genes and the 18S rRNA gene.";
RL   Mol. Biol. Evol. 16:1027-1036(1999).
DR   EMBL; AB013676; BAA83452.1; -.
DR   HSSP; P25896; 1JB0.
DR   InterPro; IPR001280; PsaA_psaB.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Chloroplast.
FT   NON_TER     604     604
SQ   SEQUENCE   604 AA;  66805 MW;  B34AE409527B0C78 CRC64;


  Query Match             71.7%;  Score 38;  DB 8;  Length 604;
  Best Local Similarity   85.7%;  Pred. No. 91;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QGNFNAW 8
           |||| ||
Db      93 QGNFEAW 99


RESULT  12
Q94BV6
ID   Q94BV6      PRELIMINARY;      PRT;   297 AA.
AC   Q94BV6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT5G27210/T21B4_120.
OS   Arabidopsis thaliana (Mouse ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Chen H., Kim C.J., Keeseman E., Meyers M.C., Banh J.,
RA   Bowser L., Carninci P., Dale J.M., Gibson H.A., Goldsmith A.D.,
RA   Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA   Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA   Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA   Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA   Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinozaki K.,
RA   Davis R.W., Theologis A., Ecker J.R.;
RT   "Arabidopsis cDNA clones.";
```

```
FL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY039860; AAK63964.1; -.
SQ   SEQUENCE   297 AA;  34133 MW;  2F2980F5B0A00F348 CRC64;

  Query Match             69.8%;  Score 37;  DB 10;  Length 297;
  Best Local Similarity   62.5%;  Pred. No. 64;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        2 QGNFNAWV 9
            ||||: |:
Db      156 QGNFSLWI 163

RESULT 13
Q9M3U4
ID   Q9M3U4      PRELIMINARY;      PRT;   345 AA.
AC   Q9M3U4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BETA 1-3 GLUCANASE.
GN   G1.
OS   Vitis vinifera (Grape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
OC   Vitis.
OX   NCBI_TaxID=29760;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. PINOT NOIR;
RA   Seibicke T., Buchholz G., Puegner A., Kassemeyer H.H.;
RT   "Investigations of PR-gene expression in grape.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ277900; CAB91554.1; -.
DR   HSSP; P12257; 1AQ0.
DR   InterPro; IPR000490; Glyco_hydro_17.
DR   Pfam; PF00332; Glyco_hydro_17; 1.
DR   PROSITE; PS00587; GLYCOSYL_HYDROL_F17; 1.
SQ   SEQUENCE   345 AA;  37456 MW;  F89DC168213ID885 CRC64;

  Query Match             69.8%;  Score 37;  DB 10;  Length 345;
  Best Local Similarity   75.0%;  Pred. No. 76;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        2 QGNFNAWV 9
            ||| |:||
Db      100 QGNANSWV 107

RESULT 14
Q94IC7
ID   Q94IC7      PRELIMINARY;      PRT;   379 AA.
AC   Q94IC7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 42.7 KDA PROTEIN (FRAGMENT).
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kimura K., Nakada Y., Itoh Y.;
RT   "Partial Arabidopsis thaliana cDNA encoding a protein of unknown
RT   function.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB062682; BAB59127.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   379 AA;  42797 MW;  C13A1D9E5B2416 CRC64;

  Query Match             69.8%;  Score 37;  DB 10;  Length 379;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4 NFNAWV 9
            ||||||
Db      117 NFNAWV 122

RESULT 15
Q07979
ID   Q07979      PRELIMINARY;      PRT;   502 AA.
AC   Q07979;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CHROMOSOME XII READING FRAME ORF YLR033W.
GN   YLR033W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Obermaier B., Piravandi E., Pinke M.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z73205; CAA97557.1; -.
DR   SGD; S0004023; YLR033W.
SQ   SEQUENCE   502 AA;  57794 MW;  8F8D01184DAE1F6F CRC64;

  Query Match             69.8%;  Score 37;  DB 3;  Length 502;
  Best Local Similarity   62.5%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 KQGNFNAW 8
            | ||:|:|
Db      283 KYGNYNSW 290

Search completed: September 5, 2002, 15:31:00
Job time: 465 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:27:33 ; Search time 58.86 Seconds
(without alignments)
3.735 Million cell updates/sec

Title: US-09-744-804-36
Perfect score: 49
Sequence: 1 NLLPPMWVT 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/7A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/8B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	49	100.0	160	2	US-08-162-402B-10	Sequence 10, Appl
2	49	100.0	487	2	US-08-162-402B-6	Sequence 6, Appl
3	49	100.0	465	2	US-08-162-402B-8	Sequence 8, Appli
4	32	65.3	22	3	US-08-940-095-163	Sequence 163, App
5	32	65.3	22	4	US-08-940-093-163	Sequence 163, App
6	32	65.3	22	2	US-08-940-096-163	Sequence 163, App
7	32	65.3	22	4	US-09-465-719-163	Sequence 163, App
8	32	65.3	22	4	US-09-453-635-163	Sequence 163, App
9	32	65.3	1147	1	US-08-131-365B-38	Sequence 38, Appl
10	32	65.3	1147	2	US-08-668-123-38	Sequence 38, Appl
11	31	63.3	342	3	US-08-785-928-1	Sequence 1, Appli
12	31	63.3	342	3	US-08-728-603-17	Sequence 17, Appl
13	30	61.2	39	4	US-08-630-915A-73	Sequence 73, Appl
14	30	61.2	40	1	US-08-318-193-6	Sequence 6, Appli
15	30	61.2	40	6	5200327 17	Patent No. 5200327
16	30	61.2	54	1	US-08-214-770-9	Sequence 9, Appli
17	30	61.2	54	5	PCT US95-02885-9	Sequence 9, Appli
18	30	61.2	852	2	US-09-070-060-3	Sequence 3, Appli
19	30	61.2	852	3	US-09-357-746-3	Sequence 3, Appli
20	30	61.2	854	2	US-09-070-060-4	Sequence 4, Appli
21	30	61.2	854	3	US-09-357-746-4	Sequence 4, Appli
22	30	61.2	1091	1	US-07-695-564-3	Sequence 3, Appli
23	30	61.2	1091	1	US-08-241-687-4	Sequence 4, Appli
24	29	59.2	50	4	US-09-177-249-262	Sequence 262, App
25	29	59.2	182	2	US-08-874-832-17	Sequence 17, Appl
26	29	59.2	182	3	US-09-097-233-17	Sequence 17, Appl
27	29	59.2	197	1	US-08-339-152A-21	Sequence 21, Appl
28	29	59.2	197	2	US-08-007-999B-12	Sequence 12, Appl
29	29	59.2	197	2	US-08-689-276A-12	Sequence 12, Appl
30	29	59.2	219	3	US-08-874-832-8	Sequence 8, Appli
31	29	59.2	219	3	US-09-097-233-8	Sequence 8, Appli
32	29	59.2	225	2	US-08-637-759B-127	Sequence 127, App
33	29	59.2	225	3	US-08-871-355A-127	Sequence 127, App
34	29	59.2	225	4	US-09-201-945-127	Sequence 127, App
35	29	59.2	409	4	US-09-258-751-419	Sequence 449, App
36	29	59.2	711	3	US-08-949-598-2	Sequence 2, Appli
37	29	59.2	781	1	US-08-373-134D-2	Sequence 2, Appli
38	29	59.2	781	2	US-09-114-637-2	Sequence 2, Appli
39	29	59.2	3033	1	US-07-925-695-5	Sequence 5, Appli
40	29	59.2	3033	1	US-07-925-695-8	Sequence 8, Appli
41	29	59.2	3033	1	US-07-925-695-9	Sequence 9, Appli
42	29	59.2	3287	2	US-08-477-451-7	Sequence 7, Appli
43	28	57.1	36	4	US-09-183-706-41	Sequence 41, Appl
44	28	57.1	36	4	US-09-567-995-41	Sequence 41, Appl
45	28	57.1	110	1	US-08-111-939-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-162-402B-10
; Sequence 10, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAPPOSA, LAVI J.
; TITLE OF INVENTION: 40 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, S brooke & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: 766 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 160 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-10

Query Match 100.0%, Score 49, DB 2, Length 160;
Best Local Similarity 100.0%, Pred. No. 0.98;

```
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           |||||||||
Db      63 NLLRRMWVT 71


RESULT   2
US-08-162-402B-6
; Sequence 6, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  387 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-6


  Query Match             100.0%;  Score 49;  DB 2;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.18;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           |||||||||
Db     131 NLLRRMWVT 139


RESULT   3
US-08-162-402B-8
; Sequence 8, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  465 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-8


  Query Match             100.0%;  Score 49;  DB 2;  Length 465;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           |||||||||
Db     209 NLLRRMWVT 217


RESULT   4
US-08-940-095-163
; Sequence 163, Application US/08940095
; Patent No. 6004925
;  GENERAL INFORMATION:
;    APPLICANT:  Dasseux, Jean-Louis
;    APPLICANT:  Sekul, Renate
;    APPLICANT:  Buttner, Klaus
;    APPLICANT:  Cornut, Isabelle
;    APPLICANT:  Metz, Gunther
;    APPLICANT:  Dufourcq, Jean
;    TITLE OF INVENTION:  APOLIPOPROTEIN A-I AGONISTS
;    TITLE OF INVENTION:  AND THEIR USE TO TREAT DYSLIPIDEMIC DISORDERS
;    NUMBER OF SEQUENCES:  258
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2811
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
```

```
;       SOFTWARE:  FastSEQ Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/940,095
;       FILING DATE:  29-SEP-1997
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Coruzzi, Laura A
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  009196-0004-999
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-493-4935
;       TELEFAX:  650-493-5556
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  163:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  22 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  No. 6004925e
US-08-940-095-163


  Query Match             65.3%;  Score 32;  DB 3;  Length 22;
  Best Local Similarity   71.4%;  Pred. No. 12;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMW 7
           ||| |:|
Db       8 NLLERLW 14


RESULT   5
US-08-940-093-163
; Sequence 163, Application US/08940093
; Patent No. 6037323
;  GENERAL INFORMATION:
;     APPLICANT:  Dasseux, Jean-Louis
;     APPLICANT:  Sekul, Renate
;     APPLICANT:  Buttner, Klaus
;     APPLICANT:  Cornut, Isabelle
;     APPLICANT:  Metz, Gunther
;     TITLE OF INVENTION:  APOLIPOPROTEIN A-I AGONISTS
;     TITLE OF INVENTION:  AND THEIR USE TO TREAT DYSLIPIDEMIC DISORDERS
;     NUMBER OF SEQUENCES:  258
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds LLP
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10036-2811
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/940,093
;       FILING DATE:  29-SEP-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Coruzzi, Laura A
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  009196-0006-999
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-493-4935
;       TELEFAX:  650-493-5556
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  163:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  22 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  No. 6037323e
US-08-940-093-163


  Query Match             65.3%;  Score 32;  DB 3;  Length 22;
  Best Local Similarity   71.4%;  Pred. No. 12;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMW 7
           ||| |:|
Db       8 NLLERLW 14


RESULT   6
US-08-940-096-163
; Sequence 163, Application US/08940096
; Patent No. 6046166
;  GENERAL INFORMATION:
;     APPLICANT:  Dasseux, Jean-Louis
;     APPLICANT:  Sekul, Renate
;     APPLICANT:  Buttner, Klaus
;     APPLICANT:  Cornut, Isabelle
;     APPLICANT:  Metz, Gunther
;     TITLE OF INVENTION:  APOLIPOPROTEIN A-I AGONISTS
;     TITLE OF INVENTION:  AND THEIR USE TO TREAT DYSLIPIDEMIC DISORDERS
;     NUMBER OF SEQUENCES:  258
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds LLP
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10036-2811
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/940,096
;       FILING DATE:  29-SEP-1997
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Coruzzi, Laura A
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  009196-0005-999
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-493-4935
;       TELEFAX:  650-493-5556
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  163:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  22 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  No. 6046166e
US-08-940-096-163


  Query Match             65.3%;  Score 32;  DB 3;  Length 22;
  Best Local Similarity   71.4%;  Pred. No. 12;
```

```
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMW 7
           ||| |:|
Db       8 NLLERLW 14


RESULT   7
US-09-465-719-163
; Sequence 163, Application US/09465719
; Patent No. 6265377
;  GENERAL INFORMATION:
;    APPLICANT:  Dasseux, Jean-Louis
;    APPLICANT:  Sekul, Renate
;    APPLICANT:  Buttner, Klaus
;    APPLICANT:  Cornut, Isabelle
;    APPLICANT:  Metz, Gunther
;    TITLE OF INVENTION:  APOLIPOPROTEIN A-1 AGONISTS
;    TITLE OF INVENTION:  AND THEIR USE TO TREAT DYSLIPIDEMIC DISORDERS
;    NUMBER OF SEQUENCES:  258
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2811
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/465,719
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/940,093
;      FILING DATE:  29-SEP-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  009196-0006-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-493-4935
;      TELEFAX:  650-493-5556
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  163:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6265377e
US-09-465-719-163


  Query Match             65.3%;  Score 32;  DB 4;  Length 22;
  Best Local Similarity   71.4%;  Pred. No. 12;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMW 7
           ||| |:|
Db       8 NLLERLW 14


RESULT   8
US-09-453-605-163
; Sequence 163, Application US/09453605
; Patent No. 6329341
;  GENERAL INFORMATION:
;       APPLICANT: Dasseux, Jean-Louis
;                  Sekul, Renate
;                  Buttner, Klaus
;                  Cornut, Isabelle
;                  Metz, Gunther
;                  Dufourcq, Jean
;       TITLE OF INVENTION: APOLIPOPROTEIN A-I AGONISTS
;                           AND THEIR USE TO TREAT DYSLIPIDEMIC DISORDERS
;       NUMBER OF SEQUENCES: 258
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Pennie & Edmonds LLP
;            STREET: 1155 Avenue of the Americas
;            CITY: New York
;            STATE: NY
;            COUNTRY: USA
;            ZIP: 10036 2811
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/453,605
;            FILING DATE: 26-No. 6329341-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/940,095
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Coruzzi, Laura A
;            REGISTRATION NUMBER: 30,742
;            REFERENCE/DOCKET NUMBER: 009196-0004-999
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650-493-4935
;            TELEFAX: 650-493-5556
;            TELEX: 66141 PENNIE
;  INFORMATION FOR SEQ ID NO: 163:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 22 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: No. 6329341e
;       SEQUENCE DESCRIPTION: SEQ ID NO: 163:
US-09-453-605-163


  Query Match             65.3%;  Score 32;  DB 4;  Length 22;
  Best Local Similarity   71.4%;  Pred. No. 12;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMW 7
           ||| |:|
Db       8 NLLERLW 14


RESULT   9
US-08-131-365B-38
; Sequence 38, Application US/08131365B
; Patent No. 5527690
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
```

```
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/131,365B
;       FILING DATE:  01-OCT-1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Parker, David L.
;       REGISTRATION NUMBER:  32,165
;       REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418-3000
;       TELEFAX:  (512) 474-7577
;   INFORMATION FOR SEQ ID NO:  38:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1147 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-131-365B-38


  Query Match             65.3%;  Score 32;  DB 1;  Length 1147;
  Best Local Similarity   62.5%;  Pred. No. 5e-02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRPMWV 8
           :|:|:||
Db     634 HLLQRLWV 641


RESULT  10
US-08-668-123-38
; Sequence 38, Application US/08668123
; Patent No. 5891631
;  GENERAL INFORMATION:
;     APPLICANT:  Brown, Michael S.
;     APPLICANT:  Briggs, Michael R.
;     APPLICANT:  Wang, Xiaodong
;     APPLICANT:  Goldstein, Joseph L.
;     TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;     TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;     TITLE OF INVENTION:  PROTEINS
;     NUMBER OF SEQUENCES:  64
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  U.S.A.
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/668,123
;       FILING DATE:  14-JUN-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/131,365
;       FILING DATE:  01-OCT-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Parker, David L.
;       REGISTRATION NUMBER:  32,165
;       REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418-3000
;       TELEFAX:  (512) 474-7577
;   INFORMATION FOR SEQ ID NO:  38:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1147 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-668-123-38


  Query Match             65.3%;  Score 32;  DB 2;  Length 1147;
  Best Local Similarity   62.5%;  Pred. No. 5e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRPMWV 8
           :|:|:||
Db     634 HLLQRLWV 641


RESULT  11
US-08-785-928-1
; Sequence 1, Application US/08785928
; Patent No. 6087115
;  GENERAL INFORMATION:
;     APPLICANT:  Gershengorn, Marvin C.
;     APPLICANT:  Arvanitakis, Leandros
;     APPLICANT:  Geras-Raaka, Elizabeth
;     APPLICANT:  Cesarman, Ethel
;     TITLE OF INVENTION:  CONSTITUTIVELY ACTIVE G PROTEIN COUPLED
;     TITLE OF INVENTION:  RECEPTOR OF HHV 8 AND METHOD OF IDENTIFYING NEGATIVE
;     TITLE OF INVENTION:  ANTAGONISTS OF G PROTEIN COUPLED RECEPTORS
;     NUMBER OF SEQUENCES:  1
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;       STREET:  Clinton Square, P.O. Box 1051
;       CITY:  Rochester
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  14603
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/785,928
;       FILING DATE:  22-JAN-1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Goldman, Michael L.
;       REGISTRATION NUMBER:  30,727
;       REFERENCE/DOCKET NUMBER:  19603/1320
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (716) 263-1304
;       TELEFAX:  (716) 263-1600
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  342 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-785-928-1


  Query Match             63.3%;  Score 31;  DB 3;  Length 342;
  Best Local Similarity   71.4%;  Pred. No. 2.4e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LLRPMWV 8
           |||| |:
Db     276 LLRRWI 282
```

```
RESULT 12
US-08-728-603-17
; Sequence 17, Application US/08728603
; Patent No. 6093806
; GENERAL INFORMATION:
;   APPLICANT: Cesarman, Ethel
;   APPLICANT: Knowles, Daniel M
;   TITLE OF INVENTION: PROTEINS OF KAPOSI'S SARCOMA ASSOCIATED
;   TITLE OF INVENTION: HERPESVIRUS
;   NUMBER OF SEQUENCES: 19
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
;     STREET: Clinton Square, P.O. Box 1051
;     CITY: Rochester
;     STATE: New York
;     COUNTRY: USA
;     ZIP: 14603
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/728,603
;     FILING DATE: 10-OCT-1996
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BRAMAN, SUSAN J.
;     REGISTRATION NUMBER: 34,103
;     REFERENCE/DOCKET NUMBER: 19603/720
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 716-263-1636
;     TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 17:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 342 amino acids
;     TYPE: amino acid
;     STRANDEDNESS:
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-728-603-17


  Query Match             63.3%;  Score 31;  DB 3;  Length 342;
  Best Local Similarity   71.4%;  Pred. No. 2.4e-02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LLRRMWV 8
          |||| |:
Db     276 LLRRWI 282


RESULT 13
US-08-630-915A-73
; Sequence 73, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
;   APPLICANT: SPARKS, Andrew B.
;   APPLICANT: HOFFMAN, No. 6309820h
;   APPLICANT: KAY, Brian K.
;   APPLICANT: FOWLKES, Dana M.
;   APPLICANT: McCONNELL, Stephen J.
;   TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
;   TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
;   TITLE OF INVENTION: USING SAME
;   NUMBER OF SEQUENCES: 227
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Pennie & Edmonds LLP
;     STREET: 1155 Avenue of the Americas
;     CITY: New York
;     STATE: New York
;     COUNTRY: USA
;     ZIP: 10036-2711
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/630,915A
;     FILING DATE: 03-APR-1996
;     CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Misrock, S. Leslie
;     REGISTRATION NUMBER: 18,872
;     REFERENCE/DOCKET NUMBER: 1101-174
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (212) 790-9090
;     TELEFAX: (212) 869-8864/9741
;     TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 73:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 39 amino acids
;     TYPE: amino acid
;     STRANDEDNESS:
;     TOPOLOGY: unknown
;   MOLECULE TYPE: peptide
US-08-630-915A-73


  Query Match             61.2%;  Score 30;  DB 4;  Length 39;
  Best Local Similarity   66.7%;  Pred. No. 45;
  Matches    6;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
            | |  ||||
Db      27 NELEDMWVT 35


RESULT 14
US-08-318-193-6
; Sequence 6, Application US/08318193
; Patent No. 5641663
; GENERAL INFORMATION:
;   APPLICANT: GARVIN, Robert T.
;   APPLICANT: MALEK, Lawrence T.
;   TITLE OF INVENTION: AN EXPRESSION SYSTEM FOR THE SECRETION
;   TITLE OF INVENTION: OF BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY
;   TITLE OF INVENTION: STIMULATING FACTOR (GM-CSF) AND OTHER HETEROLOGOUS
;   TITLE OF INVENTION: PROTEINS FROM STREPTOMYCES
;   NUMBER OF SEQUENCES: 91
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Foley & Lardner
;     STREET: 1800 Diagonal Road, Suite 500
;     CITY: Alexandria
;     STATE: Virginia
;     COUNTRY: USA
;     ZIP: 22313-0299
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/318,193
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/07/935,314
;     FILING DATE:
;     APPLICATION NUMBER: US 07/224,568
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BENT, Stephen A.
;     REGISTRATION NUMBER: 29,768
;     REFERENCE/DOCKET NUMBER: 18746/116 CACO
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  40 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-318-193-6


  Query Match             61.2%;  Score 30;  DB 1;  Length 40;
  Best Local Similarity   44.4%;  Pred. No. 46;
  Matches    4;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           |  ||:|::
Db      11 NAARRVWIS 19


RESULT  15
5200327-17
;Patent No. 5200327
;    APPLICANT: GARVIN, ROBERT T.;MALEK, LAWRENCE T.
;    TITLE OF INVENTION: EXPRESSION SYSTEM FOR THE SECRETION OF
;BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY STIMULATING
;FACTOR (GM-CSF) AND OTHER HETEROLOGOUS PROTEINS FROM
;STREPTOMYCES
;    NUMBER OF SEQUENCES: 24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/224,568
;      FILING DATE: 26-JUL-1988
;SEQ ID NO:17:
;      LENGTH: 40
5200327-17


  Query Match             61.2%;  Score 30;  DB 6;  Length 40;
  Best Local Similarity   44.4%;  Pred. No. 46;
  Matches    4;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           |  ||:|::
Db      11 NAARRVWIS 19


Search completed:  September  5, 2002, 15:27:34
Job time:  319 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:39 ; Search time 70.59 seconds
(without alignments)
10.252 Million cell updates/sec

Title: US-09-744-804-36
Perfect score: 49
Sequence: 1 NLLRRMWVT 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters. 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% query Match	Length	DB	ID	Description
1	49	100.0	387	1	MFGM_HUMAN	Q08431 homo sapien
2	44	89.8	427	1	MFGM_BOVIN	Q95114 bos taurus
3	38	77.6	224	1	PLCA_YEREN	Q56892 yersinia en
4	35	71.4	115	1	RK20_MARPO	P06385 marchantia
5	35	71.4	118	1	RK20_PINTH	P41610 pinus thunb
6	35	71.4	186	1	P152_METTH	Q26255 methanobact
7	35	71.4	409	1	MFGM_PIG	P79385 sus scrofa
8	35	71.4	909	1	QDO1_BUCAI	P57388 buchnera ap
9	34	69.4	111	1	RK20_CHLRE	P26565 chlamydomon
10	34	69.4	699	1	NU5C_DIGGR	Q32131 digitalis g
11	34	69.4	1281	1	YLR5_CAERE	P46580 caenorhabdi
12	33	67.3	234	1	ARGR_HUMAN	P55145 homo sapien
13	33	67.3	468	1	Y195_CAERE	P49049 caenorhabdi
14	32	65.3	220	1	Y4PD_RHISN	P55442 rhizobium s
15	32	65.3	354	1	GBI2_CHICK	P50147 gallus gall
16	32	65.3	390	1	YF35_HAEIN	P44246 haemophilus
17	32	65.3	394	1	VIF4_FFGRV	P14358 freq virus
18	32	65.3	445	1	EX7L_PASMU	Q9cp86 pasteurella
19	32	65.3	582	1	PGMU_POPTN	Q9asq4 populus tre
20	32	65.3	1024	1	SPE1_RAT	P56720 rattus norv
21	32	65.3	1062	1	PL70_HCMVA	P17149 human cytom
22	32	65.3	1133	1	SKE1_CRIGR	Q63416 cricetulus
23	32	65.3	1147	1	SRE1_HUMAN	P36956 homo sapien
24	32	65.3	1390	1	INSR_AEDAE	Q93105 aedes aegyp
25	32	65.3	3305	1	APLP_MANSE	Q25490 manduca sex
26	31	63.3	115	1	RK20_PORPU	P51269 porphyra pu
27	31	63.3	117	1	RK20_ARATH	P56794 arabidopsis
28	31	63.3	117	1	PL20_SYNY3	P48957 synechocyst
29	31	63.3	119	1	RK20_MAIZE	P26566 zea mays (m
30	31	63.3	127	1	RK20_SPIOL	P28903 spinacia ol
31	31	63.3	127	1	RK20_TOBAC	P06386 nicotiana t
32	31	63.3	230	1	CPSR_STRAG	Q04f62 streptococc
33	31	63.3	241	1	PL3_AQUAE	O66431 aquifex aeo
34	31	63.3	244	1	IGD1_CANFA	P49894 canis famil
35	31	63.3	244	1	PL3_AQUPY	Q9Zi50 aquifex pyr
36	31	63.3	327	1	HOLB_HAEIN	P43748 haemophilus
37	31	63.3	342	1	VG74_KSHV	Q98146 kaposi's sa
38	31	63.3	726	1	AT1_CAMPC	Q05482 camelpox vi
39	31	63.3	831	1	PRIA_SYNY3	P74397 synechocyst
40	31	63.3	887	1	GLND_KLEPN	P41393 klebsiella
41	30	61.2	118	1	PL20_THEMA	Q9x1s8 thermotoga
42	30	61.2	129	1	RPBA_BPT4	P07879 bacteriopha
43	30	61.2	146	1	PL13_MYCPN	P75178 mycoplasma
44	30	61.2	200	1	NAPC_HAEIN	P44655 haemophilus
45	30	61.2	202	1	SEGB_BPT4	P13325 bacteriopha

ALIGNMENTS

RESULT 1
MFGM_HUMAN
ID MFGM_HUMAN STANDARD; PRT; 387 AA.
AC Q08431,
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMFG)
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
GN MFGE8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast, and Breast carcinoma;
RX MEDLINE 96213908; PubMed=8639264;
RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
RT "Cloning and sequence analysis of human breast epithelial antigen
RT BA46 reveals an RGD cell adhesion sequence presented in an epidermal
RT growth factor-like domain.";
RL DNA Cell Biol. 15:281-286(1996).
RN [2]
RP SEQUENCE OF 170-387 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=91371351; PubMed=1909932;
RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
RA Ceriani R.L.;
RT "A Mr 46,000 human milk fat globule protein that is highly expressed
RT in human breast tumors contains factor VIII-like domains.";
RL Cancer Res. 51:4994-4998(1991).
RN [3]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=98194924; PubMed=9535276;
RA Giuffrida M.G., Caroletto M., Cigna C., Conti A.,
RA Godovac-Zimmermann J.;
RT "Isolation and characterization of full and truncated forms of human
RT breast carcinoma protein BA46 from human milk fat globule membranes.";
RL J. Protein Chem. 17:143-148(1998).
RN [4]
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
RX MEDLINE=99342076; PubMed=10411933;
RA Haggqvist B., Naslund J., Sletten K., Westermark G.T., Mucchiano G.,
RA Tjernberg L.O., Nordstedt C., Engstroem U., Westermark P.;
RT "Medin: an integral fragment of aortic smooth muscle cell-produced
RT lactadherin forms the most common human amyloid.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE 97405885; PubMed 9269329;
RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
RT (RGD)-dependent cell adhesion.";

```
RL   DNA Cell Biol. 16:861-869(1997).
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
CC       SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
CC   -!- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
CC       MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC   -!- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
CC       AMINO ACID 264 AND 273.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U58516; AAC50549.1; -.
DR   EMBL; S56151; AAB19771.1; -.
DR   MIM; 602281; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
FT   SIGNAL       1     23
FT   CHAIN       24    387       LACTADHERIN.
FT   CHAIN      202    387       LACTADHERIN, SHORT FORM
FT   CHAIN      268    317       MEDIN.
FT   DOMAIN      24     67       EGF-LIKE.
FT   DOMAIN      70    225       F5/8 TYPE C 1.
FT   DOMAIN     230    387       F5/8 TYPE C 2.
FT   SITE        46     48       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    27     38       BY SIMILARITY.
FT   DISULFID    32     55       BY SIMILARITY.
FT   DISULFID    57     66       BY SIMILARITY.
FT   DISULFID    70    225       BY SIMILARITY
FT   DISULFID   212    216       BY SIMILARITY.
FT   DISULFID   230    387       BY SIMILARITY.
FT   CARBOHYD   238    238       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   325    325       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   329    329       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   350    350       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   387 AA;  43123 MW;  2EE6571DEC83782D CRC64;


  Query Match          100.0%;  Score 49;  DB 1;  Length 387;
  Best Local Similarity  100.0%;  Pred. No. 0.021;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           |||||||||
Db     131 NLLRRMWVT 139


RESULT   2
MFGM_BOVIN
ID   MFGM_BOVIN     STANDARD;      PRT;   427 AA.
AC   Q95114; Q27959; P79344;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Lactadherin precursor (Milk fat globule EGF factor 8) (MFG-E8)
DE   (MGP57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
DE   SP47) (BP47) (Components 15/16).
GN   MFGE8.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC   STRAIN=HOLSTEIN; TISSUE=Mammary gland;
RX   MEDLINE=97008954; PubMed=8856064;
RA   Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RA   Petersen T.E.;
RT   "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT   milk fat globules.";
RL   Eur. J. Biochem. 240:628-636(1996).
RN   [2]
RP   SEQUENCE OF 18-427 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=96125736; PubMed=8541316;
RA   Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA   Matsuda T.;
RT   "Molecular cloning of glycoprotein antigen MGP57/53 recognized by
RT   monoclonal antibodies raised against bovine milk fat globule
RT   membrane.";
RL   Biochim. Biophys. Acta 1245:385-391(1995).
RN   [3]
RP   SEQUENCE OF 19-427 FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
RC   TISSUE=Milk;
RX   MEDLINE=93250576; PubMed=8485470;
RA   Mather I.H., Banghart L.R., Lane W.S.;
RT   "The major fat globule membrane proteins, bovine components 15/16 and
RT   guinea-pig GP 55, are homologous to MGF-E8, a murine glycoprotein
RT   containing epidermal growth factor-like and factor V/VIII like
RT   sequences.";
RL   Biochem. Mol. Biol. Int. 29:545-554(1993).
CC   -!- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
CC       MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM
CC       LACKS 53 AMINO ACIDS WITHIN THE F5/8 TYPE C 1 DOMAIN.
CC   -!- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
CC   -!- PTM: THE 2 O-LINKED GLYCANS CONSTIST OF GAL, GLCNAC AND FUC, WITH
CC       PROBABLY FUC AS REDUCING TERMINAL SUGAR.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X91895; CAA62997.1; -.
DR   EMBL; S80643; AAB35894.2; -.
DR   EMBL; Y11719; CAA72406.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
```

```
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain;
KW   Alternative splicing.
FT   SIGNAL        1     18
FT   CHAIN        19    427       LACTADHERIN.
FT   DOMAIN       20     59       EGF-LIKE 1.
FT   DOMAIN       62    106       EGF-LIKE 2.
FT   DOMAIN      109    265       F5/8 TYPE C 1.
FT   DOMAIN      270    427       F5/8 TYPE C 2.
FT   SITE         85     87       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     24     35       BY SIMILARITY.
FT   DISULFID     29     47       BY SIMILARITY.
FT   DISULFID     49     58       BY SIMILARITY.
FT   DISULFID     66     77       BY SIMILARITY.
FT   DISULFID     71     94       BY SIMILARITY.
FT   DISULFID     96    105       BY SIMILARITY.
FT   DISULFID    109    265
FT   DISULFID    252    256
FT   DISULFID    270    427
FT   CARBOHYD     27     27       O-LINKED (FUC. . .) (IN PAS-6).
FT   CARBOHYD     34     34       O-LINKED (FUC. . .) (IN PAS-7).
FT   CARBOHYD     59     59       N-LINKED (GLCNAC . .) (HYBRID) (IN PAS-6
FT                                AND PAS-7).
FT   CARBOHYD    227    227       N LINKED (GLCNAC. . .) (HIGH MANNOSE) (IN
FT                                PAS-6).
FT   VARSPLIC    169    221       MISSING (IN SHORT ISOFORM).
FT   CONFLICT     19     19       A -> F (IN REF. 1).
FT   CONFLICT     28     28       L -> Q (IN REF. 1).
SQ   SEQUENCE   427 AA;  47411 MW;  4CBBEE3A1DC4EB24 CRC64;


  Query Match             89.8%;  Score 44;  DB 1;  Length 427;
  Best Local Similarity   77.8%;  Pred. No. 0.22;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVF 9
           |:|:||||
Db     170 NIMRKMWVI 178


RESULT   3
FLGA_YEREN
ID   FLGA_YEREN     STANDARD;      PRT;   224 AA.
AC   Q56892;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Flagella basal body P-ring formation protein flgA precursor.
GN   FLGA.
OS   Yersinia enterocolitica.
OC   Bacteria; Proteobacteria; Gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=630;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W1024 / SEROTYPE O:9;
RA   Fauconnier A., Allaoui A., van Pisen A., Cornelis G., Bollen A.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN THE ASSEMBLY PROCESS OF THE P-RING
CC       FORMATION. IT MAY ASSOCIATE WITH FLGF ON THE ROD CONSTITUTING A
CC       STRUCTURE ESSENTIAL FOR THE P-RING ASSEMBLY OR MAY ACT AS A
CC       MODULATOR PROTEIN FOR THE P-RING ASSEMBLY (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Periplasmic (Probable).
CC   -!- SIMILARITY: BELONGS TO THE FLGA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z48169; CAA88191.1; -.
KW   Flagella; Periplasmic; Signal.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    224       FLAGELLA BASAL BODY P-RING FORMATION
FT                                PROTEIN FLGA.
SQ   SEQUENCE   224 AA;  24351 MW;  B08874FBC7862A49 CRC64;


  Query Match             77.6%;  Score 39;  DB 1;  Length 224;
  Best Local Similarity   62.5%;  Pred. No. 1.7;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWV 8
           |:||||:|
Db     159 NMLRRLWI 166


RESULT   4
RK20_MARPO
ID   RK20_MARPO     STANDARD;      PRT;   115 AA.
AC   P06385;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Chloroplast 50S ribosomal protein L20.
GN   RPL20.
OS   Marchantia polymorpha (Liverwort).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC   Marchantiales; Marchantiaceae; Marchantia.
OX   NCBI_TaxID=3197;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89068687; PubMed=3199436;
RA   Fukuzawa H., Kohchi T., Sano T., Shirai H., Umesono K., Inokuchi H.,
RA   Ozeki H., Ohyama K.;
RT   "Structure and organization of Marchantia polymorpha chloroplast
RT   genome. III. Gene organization of the large single copy region from
RT   rbcL to trnI(CAU).";
RL   J. Mol. Biol. 203:333-351(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA   Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA   Ozeki H.;
RT   "Chloroplast gene organization deduced from complete sequence of
RT   liverwort Marchantia polymorpha chloroplast DNA.";
RL   Nature 322:572-574(1986).
RN   [3]
RP   SEQUENCE OF 1-56 FROM N.A.
RA   Fukuzawa H., Kohchi T., Shirai H., Ohyama K., Umesono K.,
RA   Inokuchi H., Ozeki H.;
RT   "Coding sequences for chloroplast ribosomal protein S12 from the
RT   liverwort, Marchantia polymorpha, are separated far apart on the
RT   different DNA strands.";
RL   FEBS Lett. 198:11-15(1986).
CC   -!- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04465; CAA28108.1; -.
DR   EMBL; X03661; CAA27298.1; -.
DR   PIR; A02807; R5LV20.
DR   MaizeDB; 4704; MARPO_RPL20;1.
DR   InterPro; IPR001801; Ribosomal_L20.
DR   Pfam; PF00453; Ribosomal_L20; 1.
```

```
DR   PRINTS; PR00062; RIBOSOMALL20.
DR   ProDom; PD002389; Ribosomal_L20; 1
DR   PROSITE; PS00937; RIBOSOMAL_L20; 1.
KW   Ribosomal protein; rRNA-binding; Chloroplast
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   115 AA;  13475 MW;  A91B680F1D777A39 CRC64;


  Query Match             71.4%;  Score 35;  DB 1;  Length 115;
  Best Local Similarity   71.4%;  Pred. No. 3.3;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 LRRMWVT 9
           |||:|:|
Db      56 LRRLWIT 62


RESULT   5
RK20_PINTH
ID   RK20_PINTH     STANDARD;      PRT:   118 AA
AC   P41610;
DT   01-NOV 1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Chloroplast 50S ribosomal protein L20.
GN   RPL20.
OS   Pinus thunbergii (Green pine) (Japanese black pine).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus
OX   NCBI_TaxID=3350;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95024047; PubMed=7937893;
RA   Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA   Sugiura M.;
RT   "Loss of all ndh genes as determined by sequencing the entire
RT   chloroplast genome of the black pine Pinus thunbergii.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC   -!- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ----------------------------------------------------------------------
DR   EMBL; D17510; BAA04342.1; -.
DR   Mendel; 10014; PINTH;rpl20.1.
DR   InterPro; IPR001801; Ribosomal_L20.
DR   Pfam; PF00453; Ribosomal_L20; 1.
DR   PRINTS; PR00062; RIBOSOMALL20.
DR   ProDom; PD002389; Ribosomal_L20; 1.
DR   PROSITE; PS00937; RIBOSOMAL_L20; 1.
KW   Ribosomal protein; rRNA-binding; Chloroplast.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   118 AA;  13765 MW;  39FF422D658B0673 CRC64;


  Query Match             71.4%;  Score 35;  DB 1;  Length 118;
  Best Local Similarity   71.4%;  Pred. No. 3.4;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 LRRMWVT 9
           |||:|:|
Db      56 LRRLWIT 62


RESULT   6
P152_METTH
ID   P152_METTH     STANDARD;      PRT:   186 AA.
AC   Q8PZ55;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein MTH152.
GN   MTH152.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H. M., Dubois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX   MEDLINE=20473232; PubMed=11017201;
RA   Christendat D., Yee A., Dharamsi A., Kluger Y., Savchenko A.,
RA   Cort J.R., Booth V., Mackereth C.D., Saridakis V., Ekiel I.,
RA   Kozlov G., Maxwell K.L., Wu N., McIntosh L.P., Gehring K.,
RA   Kennedy M.A., Davidson A.R., Pai E.F., Gerstein M., Edwards A.M.,
RA   Arrowsmith C.H.;
RT   "Structural proteomics of an archaeon.";
RL   Nat. Struct. Biol. 7:903-909(2000).
CC   -!- COFACTOR: FMN.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SIMILARITY: BELONGS TO THE FLAVODOXIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000803; AAB84658.1; -.
DR   PDB; 1EJE; 11-OCT-00
DR   InterPro; IPR002563; Flavin_Reduct.
DR   Pfam; PF01613; Flavin_Reduct; 1.
KW   Flavoprotein; FMN; 3D-structure; Complete proteome.
SQ   SEQUENCE   186 AA;  20323 MW;  6997E8C6D9231A89 CRC64;


  Query Match             71.4%;  Score 35;  DB 1;  Length 186;
  Best Local Similarity   55.6%;  Pred. No. 5.5;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           ::: |||||
Db      81 DIIERMWVT 89


RESULT   7
MFGM_PIG
ID   MFGM_PIG       STANDARD;      PRT:   409 AA.
AC   P79385;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Lactadherin (Milk fat globule EGF factor 8) (MFG-E8) (MFGM) (Sperm
DE   surface protein SP47) (PP47).
GN   MFGE8.
```

```
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA
CC       BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC       SPERMATOZOAN
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Y11683; CAA72379.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF_like
DR   InterPro; IPR000421; FA58C
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 2
DR   SMART; SM00231; FA58C; 2
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Glycoprotein; Repeat; EGF-like domain.
FT   DOMAIN        2     41       EGF-LIKE 1.
FT   DOMAIN       44     88       EGF-LIKE 2.
FT   DOMAIN       91    247       F5/8 TYPE C 1.
FT   DOMAIN      252    409       F5/8 TYPE C 2.
FT   SITE         67     69       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID      6     17       BY SIMILARITY.
FT   DISULFID     11     29       BY SIMILARITY.
FT   DISULFID     31     40       BY SIMILARITY.
FT   DISULFID     91    247       BY SIMILARITY.
FT   DISULFID    234    238       BY SIMILARITY.
FT   DISULFID    252    409       BY SIMILARITY.
FT   CARBOHYD     41     41       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    372    372       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   409 AA;  45725 MW;  B0C07AF80029927A CRC64;


  Query Match             71.4%;  Score 35;  DB 1;  Length 409;
  Best Local Similarity   88.9%;  Pred. No. 12;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           |||||| ||
Db     152 NLLRRMPVT 160


RESULT   8
ODO1_BUCAI
ID   ODO1_BUCAI     STANDARD;      PRT;   909 AA.
AC   P57388;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   2 oxoglutarate dehydrogenase E1 component (EC 1.2.4.2) (Alpha-
DE   ketoglutarate dehydrogenase).
GN   SUCA OR BU302.
OS   Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS   symbiotic bacterium).
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX   NCBI_TaxID=118099;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN: TOKYO 1998;
RX   MEDLINE=20445173; PubMed=10993077;
RA   Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT   "Genome sequence of the endocellular bacterial symbiont of aphids
RT   Buchnera sp. APS.";
RL   Nature 407:81-86(2000).
CC   -!- FUNCTION: THE 2-OXOGLUTARATE DEHYDROGENASE COMPLEX CATALYZES THE
CC       OVERALL CONVERSION OF 2-OXOGLUTARATE TO SUCCINYL-COA & CO(2). IT
CC       CONTAINS MULTIPLE COPIES OF 3 ENZYMATIC COMPONENTS: 2 OXOGLUTARATE
CC       DEHYDROGENASE (E1), DIHYDROLIPOAMIDE SUCCINYLTRANSFERASE (E2) AND
CC       LIPOAMIDE DEHYDROGENASE (E3) (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 2-oxoglutarate + lipoamide = S-
CC       succinyldihydrolipoamide + CO(2).
CC   -!- COFACTOR: THIAMINE PYROPHOSPHATE (BY SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP001118; BAB13011.1; -.
DR   InterPro; IPR001017; E1_dh.
DR   InterPro; IPR002208; PPTA.
DR   Pfam; PF00676; E1_dehydrog; 1.
KW   Glycolysis; Oxidoreductase; Flavoprotein; Thiamine pyrophosphate;
KW   Complete proteome.
SQ   SEQUENCE   909 AA;  105836 MW;  AE749F8CD9F9CF8B CRC64;


  Query Match             71.4%;  Score 35;  DB 1;  Length 909;
  Best Local Similarity   55.6%;  Pred. No. 29;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           ||| : |:|
Db     157 NLLEKQWIT 165


RESULT   9
RK20_CHLRE
ID   RK20_CHLRE     STANDARD;      PRT;   111 AA.
AC   P26565;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Chloroplast 50S ribosomal protein L20.
GN   RPL20.
OS   Chlamydomonas reinhardtii.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2137;
RA   Yu W., Zhang D., Spreitzer R.J.;
RT   "Sequences of the Chlamydomonas reinhardtii chloroplast genes encoding
RT   tRNA Ser and ribosomal protein L20.";
RL   Plant Physiol. 100:1079-1080(1992).
CC   -!- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X62566; CAA44439.1; -.
DR   PIR; S18026; R5YM20.
DR   Mendel; 4049; CHLre;rpl20;1.
DR   InterPro; IPR001081; Ribosomal_L20.
DR   Pfam; PF00453; Ribosomal_L20; 1.
DR   PRINTS; PR00062; RIBOSOMALL20.
DR   ProDom; PD002389; Ribosomal_L20; 1.
DR   PROSITE; PS00937; RIBOSOMAL_L20; 1.
KW   Ribosomal protein; rRNA-binding; Chloroplast.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   111 AA;  13417 MW;  62C4F729BF15FBB9 CRC64;


  Query Match             69.4%;  Score 34;  DB 1;  Length 111;
  Best Local Similarity   83.3%;  Pred. No. 5;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 RRMWVT 9
           ||||:|
Db      57 RRMWIT 62


RESULT  10
NU5C_DIGGR
ID   NU5C_DIGGR     STANDARD;      PRT;   699 AA.
AC   Q32131;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   NADH plastoquinone oxidoreductase chain 5, chloroplast (EC 1.6.5.3)
DE   (Fragment).
GN   NDHF.
OS   Digitalis grandiflora (Yellow foxglove).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Lamiales; Veronicaceae; Digitalis.
OX   NCBI_TaxID=38791;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Olmstead R.G., Reeves P.A.;
RT   "Evidence for the polyphyly of the Scrophulariaceae based on
RT   chloroplast rbcL and ndhF sequences.";
RL   Ann. Mo. Bot. Gard. 82:176-193(1995).
CC   -!- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as it's content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L36499; AAA84203.1; -.
DR   Mendel; 2414; DIGqr;ndhF;1.
DR   InterPro; IPR001750; Oxidored_q1
DR   InterPro; IPR002128; Oxidored_q1_C.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF01010; oxidored_q1_C; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
KW   Oxidoreductase; NAD; Plastoquinone; Chloroplast.
FT   NON_TER       1      1
FT   NON_TER     699    699
SQ   SEQUENCE   699 AA;  79309 MW;  5685FBA561C63D01 CRC64;


  Query Match             69.4%;  Score 34;  DB 1;  Length 699;
  Best Local Similarity   85.7%;  Pred. No. 34;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMW 7
           | |||||
Db      26 NKLRRMW 32


RESULT  11
YLB5_CAEEL
ID   YLB5_CAEEL     STANDARD;      PRT;  1281 AA.
AC   P46580;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 146.8 kDa protein C34E10.5 in chromosome III.
GN   C34E10.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Kirsten J.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC   -!- SIMILARITY: TO S.POMBE SKB1 AND YEAST YBR133C.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U10402; AAA19067.1; -.
DR   WormPep; C34E10.5; CE01185.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Transmembrane; Zinc-finger.
FT   TRANSMEM     69     89       POTENTIAL.
FT   TRANSMEM   1165   1185       POTENTIAL.
FT   ZN_FING    1222   1267       RING-TYPE.
SQ   SEQUENCE   1281 AA;  146??? MW;  7D27275ABB540D04 CRC64;


  Query Match             69.4%;  Score 34;  DB 1;  Length 1281;
  Best Local Similarity   55.6%;  Pred. No. 65;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           |||: :|: |
Db     271 NLLKHLWTT 279


RESULT  12
ARGR_HUMAN
ID   ARGR_HUMAN     STANDARD;      PRT;   234 AA.
AC   P55145;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Arginine-rich protein.
GN   ARP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANT CANCER ARG-50.
RX   MEDLINE=96211400; PubMed=8649854;
RA   Shridhar V., Rivard S., Shridhar R., Mullins C., Bostick D., Saki W.,
RA   Grignon D., Miller O.J., Smith D.I.;
RT   "A gene from human chromosomal band 3p21.1 encodes a highly conserved
RT   arginine-rich protein and is mutated in renal cell carcinomas.",
RL   Oncogene 12:1931-1939(1996).
RN   [2]
RP   VARIANT ARG-50.
RX   MEDLINE=97136272; PubMed=8971156;
RA   Shridhar R., Shridhar V., Rivard S., Siegfried J.M.,
RA   Pietraszkiewicz H., Ensley J., Pauley R., Grignon D., Saki W.,
RA   Miller O.J., Smith D.I.;
RT   "Mutations in the arginine-rich protein gene, in lung, breast, and
RT   prostate cancers, and in squamous cell carcinoma of the head and
RT   neck.";
RL   Cancer Res. 56:5576-5578(1996).
RN   [3]
RP   VARIANTS IN PANCREATIC CANCERS.
RX   MEDLINE=97316783; PubMed=9174057;
RA   Shridhar V., Rivard S., Wang X., Shridhar R., Paisley C., Mullins C.,
RA   Beirnat I., Dugan M., Sarkar F., Miller O.J., Vaitkevicius V.K.,
RA   Smith D.I.;
RT   "Mutations in the arginine-rich protein gene (ARP) in pancreatic
RT   cancer.";
RL   Oncogene 14:2213-2216(1997).
CC   -!- SIMILARITY: BELONGS TO THE ARP FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M83751; AAB08753.1; ALT_INIT.
DR   MIM; 601916; -
KW   Disease mutation.
FT   DOMAIN       38      55       POLY-ARG.
FT   VARIANT      50      50       M -> R (IN CANCER).
FT                                 /FTId=VAR_010245.
FT   VARIANT      50      50       MISSING (IN CANCER).
FT                                 /FTId=VAR_010246
FT   VARIANT      51      51       R -> K (IN CANCER).
FT                                 /FTId=VAR_010247
SQ   SEQUENCE   234 AA;  26906 MW;  B8AB0E7E8C079C94 CRC64;


  Query Match             67.3%;  Score 33;  DB 1;  Length 234;
  Best Local Similarity   71.4%;  Pred. No. 17;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 LRRMWVT 9
           :|||| |
Db      53 MRRMWAT 59


RESULT  13
YJ95_CAEEL
ID   YJ95_CAEEL     STANDARD;      PRT;   468 AA.
AC   P49049;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 52.8 kDa protein T05E11.5 in chromosome IV.
GN   T05E11.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Kershaw J.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO YEAST YKL100C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z68751; CAA92975.1; -.
DR   WormPep; T05E11.5; CE06364.
KW   Hypothetical protein.
SQ   SEQUENCE   468 AA;  52793 MW;  00DFFACB679AA5F8 CRC64;


  Query Match             67.3%;  Score 33;  DB 1;  Length 468;
  Best Local Similarity   55.6%;  Pred. No. 36;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 LRRMWVT 9
           :|||:| |:|
Db     263 HLLKKHWIT 271


RESULT  14
Y4FD_RHISN
ID   Y4FD_RHISN     STANDARD;      PRT;   220 AA.
AC   P55442;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 24.6 kDa protein Y4FD.
GN   Y4FD.
OS   Rhizobium sp. (strain NGR234).
OG   Plasmid sym pNGR234a.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=394;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97385956; PubMed=9163424;
RA   Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA   Perret X.;
RT   "Molecular basis of symbiosis between Rhizobium and legumes.";
RL   Nature 387:394-401(1997).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: NONE OBVIOUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000072; AAB91661.1; -.
KW   Hypothetical protein; Transmembrane; Plasmid.
FT   TRANSMEM      9     29       POTENTIAL.
FT   TRANSMEM     54     74       POTENTIAL.
FT   TRANSMEM    105    125       POTENTIAL.
FT   TRANSMEM    177    197       POTENTIAL.
SQ   SEQUENCE   220 AA;  24627 MW;  50411A4A84557381 CRC64;


  Query Match             67.3%;  Score 33;  DB 1;  Length 220;
```

```
  Best Local Similarity   57.1%;  Pred. No. 25;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 LRRMWVT 9
           |::|:|
Db       6 LRKLWIT 12


RESULT  15
GBI2_CHICK
ID   GBI2_CHICK     STANDARD;      PRT;   354 AA.
AC   P50147;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Guanine nucleotide-binding protein G(i), alpha-2 subunit (Adenylate
DE   cyclase-inhibiting G alpha protein).
GN   GNAI2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95121926; PubMed=7821803;
RA   Kilbourne E.J., Galper J.B.;
RT   "Cloning of cDNAs coding for the G alpha i1 and G alpha i2 G-proteins
RT   from chick brain.";
RL   Gene 150:341-344(1994).
CC   -!- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
CC       INVOLVED AS MODULATORS OR TRANSDUCERS IN VARIOUS TRANSMEMBRANE
CC       SIGNALING SYSTEMS.
CC   -!- FUNCTION: THE G(I) PROTEINS ARE INVOLVED IN HORMONAL REGULATION OF
CC       ADENYLATE CYCLASE: THEY INHIBIT THE CYCLASE IN RESPONSE TO BETA-
CC       ADRENERGIC STIMULI.
CC   -!- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC       THE ALPHA CHAIN CONTAINS THE GUANINE NUCLEOTIDE BINDING SITE.
CC   -!- SIMILARITY: BELONGS TO THE G-ALPHA FAMILY. SUBFAMILY 1
CC       (G(I/O/T/Z)).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L24549; AAA65067.1; -.
DR   HSSP; P10824; 1AS3.
DR   InterPro; IPR001019; Gprotein_alpha.
DR   InterPro; IPR001230; Prenyltn.
DR   Pfam; PF00503; G-alpha; 1.
DR   PRINTS; PR00318; GPROTEINA.
DR   SMART; SM00275; G-alpha; 1.
KW   GTP-binding; Transducer; ADP-ribosylation; Multigene family;
KW   Myristate; Palmitate; Lipoprotein.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   LIPID         1      1       MYRISTATE (BY SIMILARITY).
FT   LIPID         2      2       PALMITATE (BY SIMILARITY).
FT   NP_BIND      39     46       GTP (BY SIMILARITY).
FT   NP_BIND     200    204       GTP (BY SIMILARITY).
FT   NP_BIND     269    272       GTP (BY SIMILARITY).
FT   MOD_RES     178    178       ADP-RIBOSYL[1] (BY ACTION OF CTX).
FT   MOD_RES     351    351       ADP-RIBOSYL[1] (BY ACTION OF IAP).
SQ   SEQUENCE   354 AA;  40446 MW;  D9645493D9C5CC4F CRC64;


  Query Match             65.3%;  Score 32;  DB 1;  Length 354;
  Best Local Similarity   57.1%;  Pred. No. 42;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMW 7
             |::||:|
Db     125 NVIRRLW 131


Search completed: September  5, 2002, 15:31:40
Job time: 475 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:28:43 ; Search time 69.78 Seconds
(without alignments)
12.393 Million cell updates/sec

Title: US-09-744-804-36
Perfect score: 49
Sequence: 1 NLLRPMWVT 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	89.8	401	2	S65138	glycoprotein antig
2	44	89.8	427	2	S74211	PAS 6/7 protein pr
3	38	77.6	224	2	S54218	flgA protein - Yer
4	36	73.5	232	2	AC0219	flagella basal bod
5	36	73.5	390	2	T51197	hypothetical prote
6	36	73.5	405	2	AI3148	conserved hypothet
7	36	73.5	425	2	B98139	bmo3 protein (AF07
8	36	73.5	3071	2	T45584	hypothetical prote
9	35	71.4	116	1	R5EV20	ribosomal protein
10	35	71.4	119	2	T07464	ribosomal protein
11	35	71.4	186	1	G69069	hypothetical prote
12	35	71.4	409	2	T11743	pP47 protein - pig
13	35	71.4	909	2	C84965	oxoglutarate dehyd
14	34	69.4	112	1	R5KM20	ribosomal protein
15	34	69.4	338	2	T24743	hypothetical prote
16	34	69.4	645	2	T12159	NADH dehydrogenase
17	34	69.4	660	2	E95084	cation-transportin
18	34	69.4	707	2	T12658	NADH dehydrogenase
19	34	69.4	707	2	T12664	NADH dehydrogenase
20	34	69.4	707	2	T12665	NADH dehydrogenase
21	34	69.4	707	2	T12668	NADH dehydrogenase
22	34	69.4	707	2	T12671	NADH dehydrogenase
23	34	69.4	707	2	T12751	NADH dehydrogenase
24	34	69.4	707	2	T12752	NADH dehydrogenase
25	34	69.4	707	2	T12759	NADH dehydrogenase
26	34	69.4	707	2	T13032	NADH dehydrogenase
27	34	69.4	707	2	T13036	NADH dehydrogenase
28	34	69.4	707	2	T34444	hypothetical prote
29	34	69.4	733	2	S76856	hypothetical prote
30	34	69.4	750	2	A97952	H+/K+-exchanging A
31	34	69.4	1281	2	T15762	hypothetical prote
32	33	67.3	71	2	G83792	hypothetical prote
33	33	67.3	234	2	S27956	arginine-rich prot
34	33	67.3	334	2	C70813	hypothetical prote
35	33	67.3	416	2	AE2527	hypothetical prote
36	33	67.3	468	2	T24523	hypothetical prote
37	[illegible]	[illegible]	[illegible]	2	[illegible]	[illegible] GDEF
38	32	65.3	153	1	FQMSGM	granulocyte-macrop
39	32	65.3	247	2	A57266	probable methane m
40	32	65.3	294	2	B70692	hypothetical prote
41	32	65.3	294	2	T45662	hypothetical prote
42	32	65.3	307	2	AD2179	hypothetical prote
43	32	65.3	343	2	C84711	hypothetical prote
44	32	65.3	345	2	T19834	hypothetical prote
45	32	65.3	355	2	I50238	C12 protein alpha-

ALIGNMENTS

RESULT 1
S65138
glycoprotein antigen MGP57/53, mammary gland - bovine (fragment)
N;Alternate names: glycoprotein component 15/major fat globule membrane protein 15/MFG-E
C;Species: Bos primigenius taurus (cattle)
C;Date: 28 Oct 1996 #sequence_revision 13 Mar 1997 #text_change 07 Aug 1998
C;Accession: S65138; G48394
R;Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A;Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal
A;Reference number: S65138; MUID:96125736
A;Accession: S65138
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-401 <AOK>
R;Mather, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A;Title: The major fat globule membrane proteins, bovine components 15/16 and guinea-II-like sequences.
A;Reference number: A48394; MUID:93250576
A;Accession: G48394
A;Status: preliminary
A;Molecule type: protein
A;Residues: 207-220 <MAT>
A;Experimental source: milk
A;Note: sequence extracted from NCBI backbone (NCBIP:131457)
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
C;Keywords: glycoprotein
F;1-22/Domain: EGF homology (fragment) <EG1>
F;19-79/Domain: EGF homology <EG2>
F;82-239/Domain: discoidin I amino-terminal homology <DN1>
F;243-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match 89.8%; Score 44; DB 2; Length 401;
Best Local Similarity 77.8%; Pred. No. 0.69;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 NLLRPMWVT 9
                ||:|:||||
Db        144 NLMRKMWVT 152
```

RESULT 2
S74211
PAS-6/7 protein precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 21-Jan-2000
C;Accession: S74211; S78114; S24181
R;Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
Eur. J. Biochem. 240, 628-636, 1996
A;Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat g

A;Reference number: S74211; MUID:97008954
A;Accession: S74211
A;Molecule type: mRNA
A;Residues: 1-427 <HVA>
A;Cross-references: EMBL:X91895, NID:g1632778, PIDN:CAA62997.1, PID:g1632779
A;Accession: S78114
A;Molecule type: protein
A;Residues: 19-85,96-120,140-165,174-[illegible],221-232,248-277,285-[illegible],309-337,339-420,425-427
R;Kim, D.H.; Kanno, C.; Mizokami, Y.
Biochim. Biophys. Acta 1122, 203-211, 1992
A;Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from
A;Reference number: S23926; MUID:92353107
A;Accession: S24181
A;Molecule type: protein
A;Residues: 383-394 <KIM>
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
C;Keywords: blocked amino end; disulfide bond; glycoprotein, milk
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-427/Product: PAS 6/7 protein #status experimental <MAT>
F;24-58/Domain: EGF homology <EG1>
F;66-105/Domain: EGF homology <EG2>
F;108-265/Domain: discoidin I amino terminal homology <DN1>
F;269-427/Domain: discoidin I amino terminal homology <DN2>
F;24-35,29-47,49-58,66-77,71-94,96-105/Disulfide bonds: #status predicted
F;27/Binding site: carbohydrate (Ser) (covalent) #status experimental
F;34/Binding site: carbohydrate (Thr) (covalent) #status experimental
F;59,227/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;109-265,252-256,270-427/Disulfide bonds: #status experimental

Query Match 89.8%; Score 44; DB 2; Length 427;
Best Local Similarity 77.8%; Pred. No. 0.74;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLLRRMWVT 9
|||:|||||
Db 170 NLMRKMWVT 178

RESULT 3
S54218
flgA protein - Yersinia enterocolitica
C;Species: Yersinia enterocolitica
C;Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 29-Sep-1999
C;Accession: S54218
R;Fauconnier, A.; Allaoui, A.; van Elsen, A.; Cornelis, G.; Bollen, A.
submitted to the EMBL Data Library, February 1995
A;Description: Clustering of flagellar genes around invA, the Yersinia enterocolitica in
A;Reference number: S54213
A;Accession: S54218
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-224 <FAU>
A;Cross-references: EMBL:X48169; NID:g793891; PIDN:CAA88191.1; PID:g793897
C;Superfamily: flagellar basal body P-ring protein flgA

Query Match 77.6%; Score 38; DB 2; Length 224;
Best Local Similarity 62.5%; Pred. No. 5.7;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLLRRMWV 8
|:|||:|:
Db 159 NMLRRLWI 166

RESULT 4
AC0219
flagella basal body P-ring formation protein FlgA [imported] - Yersinia pestis (strain C
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C;Accession: AC0219
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Varga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0031; MUID:21470413; PMID:11586360
A;Accession: AC0219
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-232 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC90615.1; PID:g15979821; GSPDB:GN00175
C;Genetics:
A;Gene: flgA
C;Superfamily: flagellar basal body P-ring protein flgA

Query Match 77.6%; Score 38; DB 2; Length 232;
Best Local Similarity 62.5%; Pred. No. 15;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLLRRMWV 8
|:|||:|:
Db 167 NMLRRLWL 174

RESULT 5
T51197
hypothetical protein B7N4.40 [imported] Neurospora crassa
C;Species: Neurospora crassa
C;Date: 28-Jul-2000 #sequence_revision 28-Jul-2000 #text_change 08-Sep-2000
C;Accession: T51197
R;Schulte, U.; Aign, V.; Boheisel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, July 2000
A;Reference number: Z25286
A;Accession: T51197
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-390 <SCH>
A;Cross-references: EMBL:AL390218; GSPDB:GN00316; NCSP:B7N4.40
A;Experimental source: BAC clone B7N4; strain OR74A
C;Genetics:
A;Gene: NCSP:B7N4.40
A;Map position: 6
A;Introns: 56/3
C;Superfamily: Neurospora crassa hypothetical protein B7N4.40

Query Match 73.5%; Score 36; DB 2; Length 390;
Best Local Similarity 85.7%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLLRRMW 7
| |||||
Db 272 NFLRRMW 278

RESULT 6
AI3148
conserved hypothetical protein bme3 [imported] - Agrobacterium tumefaciens (strain C5
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AI3148
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McC
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon Kam
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58
A;Reference number: AB2577; PMID:11743193
A;Accession: AI3148
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-405 <KUR>

```
A;Cross-references: GB:AE008689; PIDN:AAL45607.1; PID:g17743427; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: bme3
A;Map position: linear chromosome

  Query Match             73.5%;  Score 36;  DB 2;  Length 405;
  Best Local Similarity   75.0%;  Pred. No. 25;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LLRRMWVT 9
              |:||:|||
Db      328 LVRRIWVT 335

RESULT   7
B98139
bme3 protein (AE007290) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: B98139
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: B98139
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-425 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK88636.1; PID:g15158357; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_143
A;Map position: linear chromosome

  Query Match             73.5%;  Score 36;  DB 2;  Length 425;
  Best Local Similarity   75.0%;  Pred. No. 26;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LLRRMWVT 9
              |:||:|||
Db      348 LVRRIWVT 355

RESULT   8
T45584
hypothetical protein F11C1.220 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 18-Feb-2000
C;Accession: T45584
R;Bargues, M.; Collado, M.C.; Navarro, P.; Terol, J.; Perez-Alonso, M.; Mewes, H.W.; May
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z23007
A;Accession: T45584
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3071 <BAR>
A;Cross-references: EMBL:AL132976
A;Experimental source: cultivar Columbia, BAC clone F11C1
C;Genetics:
A;Map position: 3
A;Introns: [illegible]
419 [illegible]
A;Note: F11C1.220
C;Superfamily: Arabidopsis thaliana hypothetical protein F11C1.220

  Query Match             73.5%;  Score 36;  DB 2;  Length 3071;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 NLLRRMWV 8
              |||| ||:
Db      601 NLLRRMWL 608

RESULT   9
R5LV20
ribosomal protein L20 - liverwort (Marchantia polymorpha) chloroplast
C;Species: chloroplast Marchantia polymorpha
C;Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 22-Jun-1999
C;Accession: A02807; S01547
R;Ohyama, K.
submitted to the EMBL Data Library, October 1986
A;Reference number: A00150
A;Accession: A02807
A;Molecule type: DNA
A;Residues: 1-116 <OHY>
R;Ohyama, K.; Fukuzawa, H.; Kohchi, T.; Shirai, H.; Sano, T.; Sano, S.; Umesono, K.;
Nature 322, 572-574, 1986
A;Title: Chloroplast gene organization deduced from complete sequence of liverwort Ma
A;Reference number: A38014
A;Contents: annotation; gene organization, sites, features
R;Fukuzawa, H.; Kohchi, T.; Sano, T.; Shirai, H.; Umesono, K.; Inokuchi, H.; Ozeki, H
J. Mol. Biol. 203, 333-351, 1988
A;Title: Structure and organization of Marchantia polymorpha chloroplast genome. III.
A;Reference number: S01529; MUID:89068687
A;Accession: S01547
A;Molecule type: DNA
A;Residues: 1-116 <FUK>
A;Cross-references: GB:X04465; GB:Y00686; NID:g11646; PIDN:CAA28108.1; PID:g11696
C;Genetics:
A;Gene: rpl20
A;Genome: chloroplast
C;Superfamily: Escherichia coli ribosomal protein L20
C;Keywords: chloroplast; protein biosynthesis; ribosome

  Query Match             71.4%;  Score 35;  DB 1;  Length 116;
  Best Local Similarity   71.4%;  Pred. No. 12;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 LRRMWVT 9
              |||:|:|
Db       57 LRRLWIT 63

RESULT   10
T07464
ribosomal protein L20 - Japanese black pine chloroplast
C;Species: chloroplast Pinus thunbergiana (Japanese black pine)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 18-Aug-2000
C;Accession: T07464
R;Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugiura, M.
Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994
A;Title: Loss of all ndh genes as determined by sequencing the entire chloroplast gen
A;Reference number: Z16030; MUID:95024047
A;Accession: T07464
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-119 <WAK>
A;Cross-references: EMBL:D17510; NID:g529643; PIDN:BAA04342.1; PID:g1262625
C;Genetics:
A;Gene: rpl20
A;Genome: chloroplast
C;Superfamily: Escherichia coli ribosomal protein L20
C;Keywords: chloroplast; ribosome

  Query Match             71.4%;  Score 35;  DB 2;  Length 119;
  Best Local Similarity   71.4%;  Pred. No. 12;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 LRRMWVT 9
```

```
          |||:|:|
Db      57 LRRIWIT 63


RESULT  11
G69069
hypothetical protein MTH152 - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: G69069
R;Smith, D.R.; Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T.,
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J., Gibson, R.; Jiwani, N.
Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514
A;Accession: G69069
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-186 <MTH>
A;Cross-references: GB:AE000803; GB:AE000666; NID:g2621179; PIDN:AAB84658.1; PID:g2621179
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH152
A;Start codon: GTG
C;Superfamily: Bacillus subtilis hypothetical protein ywrF


  Query Match             71.4%;  Score 35;  DB 1;  Length 186;
  Best Local Similarity   55.6%;  Pred. No. 18;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           ::: |||||
Db      81 DIIFRMWVT 89


RESULT  12
T11743
pP47 protein - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T11743
R;Ensslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H., Schmidtke, J., Matsuda, T., Toepfe
Biol. Reprod. 58, 1057-1064, 1998
A;Title: Molecular cloning and characterization of P47, a novel boar sperm-associated zo
A;Reference number: Z17325; MUID:98206817
A;Accession: T11743
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-409 <ENS>
A;Cross-references: EMBL:Y11683, NID:g2652927, PIDN:CAA72379.1, PID:g2652928
A;Experimental source: testis
C;Function:
A;Description: may be involved in membrane remodeling and/or function as a zona pellucid
C;Superfamily: unassigned EGF-related proteins; EGF homology
F;6-40/Domain: EGF homology <EGF>


  Query Match             71.4%;  Score 35;  DB 2;  Length 409;
  Best Local Similarity   88.9%;  Pred. No. 40;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           |||||| ||
Db     152 NLLRRMRVT 160


RESULT  13
C84965
oxoglutarate dehydrogenase (lipoamide) (EC 1.2.4.2) e1 component [imported] - Buchnera s
C;Species: Buchnera sp.
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: C84965
R;Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A;Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp
A;Reference number: A84930; MUID:20445173
A;Accession: C84965
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-909 <STO>
A;Cross-references: GB:AP000398; GSPDB:GN00144
A;Experimental source: strain APS
C;Genetics:
A;Gene: sucA; BU302
C;Superfamily: oxoglutarate dehydrogenase (lipoamide); thiamin pyrophosphate-binding
C;Keywords: oxidoreductase


  Query Match             71.4%;  Score 35;  DB 2;  Length 909;
  Best Local Similarity   55.6%;  Pred. No. 87;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           ||| : |:|
Db     157 NLLEKQWIT 165


RESULT  14
R5KM20
ribosomal protein L20, chloroplast - Chlamydomonas reinhardtii chloroplast
C;Species: chloroplast Chlamydomonas reinhardtii
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 22-Jun-1999
C;Accession: S18026
R;Yu, W.; Zhang, D.; Spreitzer, R.J.
submitted to the EMBL Data Library, October 1991
A;Description: Sequences of the trnS and rpl20 genes of the Chlamydomonas reinhardtii
A;Reference number: S18026
A;Accession: S18026
A;Molecule type: DNA
A;Residues: 1-112 <YUW>
A;Cross-references: EMBL:X62566, NID:g11463, PIDN:CAA44439.1; PID:g11464
A;Experimental source: strain 2137 mt+
C;Genetics:
A;Gene: rpl20
A;Genome: chloroplast
C;Superfamily: Escherichia coli ribosomal protein L20
C;Keywords: chloroplast; protein biosynthesis; ribosome


  Query Match             69.4%;  Score 34;  DB 1;  Length 112;
  Best Local Similarity   83.3%;  Pred. No. 17;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 RRMWVT 9
           ||||:|
Db      58 RPMWIT 63


RESULT  15
T24743
hypothetical protein T09E11.10 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jan-2000
C;Accession: T24743
R;McLay, K.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19930
A;Accession: T24743
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-338 <WIL>
A;Cross-references: EMBL:Z81147, PIDN:CAB03547.1, GSPDB:GN00019, CESP:T09E11.10
A;Experimental source: clone T09E11
C;Genetics:
```

A;Gene: CESP:T09E11.10
A;Map position: 1
A [illegible]
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

Query Match 69.4%; Score 34; DB 2; Length 338;
Best Local Similarity 55.6%; Pred. No. 51;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLLRRMWVT 9
|:||: |:|
Db 96 NVLRKTWMT 104

Search completed: September 5, 2002, 15:28:50
Job time: 355 sec

GenCore version 4.5

QM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:09 ; Search time 122.86 Seconds
(without alignments)
12.673 Million cell updates/sec

Title: US-09-744-804-36
Perfect score: 49
Sequence: 1 NLLRFMWVT 9

Scoring table: PIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distributions.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	49	100.0	335	4	Q9BTL9	Q9btl9 homo sapien
2	41	83.7	1928	4	Q99736	Q99736 homo sapien
3	41	83.7	2392	4	O95001	O95001 homo sapien
4	41	83.7	2412	4	Q92616	Q92616 homo sapien
5	36	73.5	390	3	Q9P3C0	Q9p3c0 neurospora
6	36	73.5	3071	10	Q9SND0	Q9snd0 arabidopsis
7	35	71.4	1292	10	Q91Q50	Q91q50 oryza sativ
8	34	69.4	232	5	Q25913	Q25913 plasmodium
9	34	69.4	329	4	Q96L58	Q96l58 homo sapien
10	34	69.4	338	5	O02310	O02310 caenorhabdi
11	34	69.4	409	2	Q93CU8	Q93cu8 shigella bo
12	34	69.4	645	8	O46945	O46945 heliotropiu
13	34	69.4	660	16	Q97RR4	Q97rr4 streptococc
14	34	69.4	700	8	Q9TIT9	Q9tit9 tournefort
15	34	69.4	706	8	Q9TL48	Q9tl48 rhamnus dav
16	34	69.4	707	5	Q23034	Q23034 caenorhabdi
17	34	69.4	707	8	P92256	P92256 ceanothus a
18	34	69.4	707	8	P92261	P92261 ceanothus c
19	34	69.4	707	8	P92265	P92265 ceanothus f
20	34	69.4	707	8	P92268	P92268 ceanothus i
21	34	69.4	707	8	P92279	P92279 ceanothus s
22	34	69.4	707	8	P92281	P92281 ceanothus t
23	34	69.4	707	8	P92284	P92284 ceanothus v
24	34	69.4	707	8	P92262	P92262 ceanothus c
25	34	69.4	707	8	P92272	P92272 ceanothus p
26	34	69.4	707	8	P92273	P92273 ceanothus p
27	34	69.4	733	16	P74652	P74652 synechocyst
28	34	69.4	746	8	Q9TLB7	Q9tlb7 digitalis p
29	34	69.4	884	10	Q941I1	Q941i1 oryza sativ
30	34	69.4	983	10	Q9LI34	Q9li34 oryza sativ
31	33	67.3	71	16	Q9KDP8	Q9kdp8 bacillus ha
32	33	67.3	162	2	Q9KBL4	Q9kbl4 streptomyce
33	33	67.3	182	4	Q96IS4	Q96is4 homo sapien
34	33	67.3	237	16	Q92RC0	Q92rc0 rhizobium m
35	33	67.3	331	16	O53855	O53855 mycobacteri
36	33	67.3	431	8	Q9MS91	Q9ms91 ascarina lu
37	33	67.3	475	10	Q9LJY1	Q9ljy1 arabidopsis
38	33	67.3	547	10	Q9LH23	Q9lh23 oryza sativ
39	33	67.3	700	8	Q9TLG0	Q9tlg0 amphianthus
40	33	67.3	700	8	Q9TI66	Q9ti66 gratiola pi
41	33	67.3	736	8	Q9TJS1	Q9tjs1 pauridianth
42	33	67.3	805	16	Q9RX49	Q9rx49 deinococcus
43	33	67.3	956	5	Q9NNG5	Q9nng5 leishmania
44	33	67.3	1076	5	Q9W2P1	Q9w2p1 drosophila
45	33	67.3	4010	5	Q9VJC6	Q9vjc6 drosophila

ALIGNMENTS

RESULT 1
Q9BTL9
ID Q9BTL9 PRELIMINARY; PRT; 335 AA.
AC Q9BTL9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO MILK FAT GLOBULE-EGF FACTOR 8 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA.;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 1 EGF/8 TYPE C DOMAIN.
DR EMBL; BC003610; AAH03610.1; -.
DR HSSP; P08709; 1BF9.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR000421; FA58_C.
DR PRINTS; PR00010; EGFBLOOD.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01285; FA58C_1; 1.
DR PROSITE; PS01286; FA58C_2; 2.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 335 AA; 37523 MW; 20E84E5DB78E8E25 CRC64;

Query Match 100.0%; Score 49; DB 4; Length 335;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NLLRRMWVT 9
           |||||||||
Db     131 NLLRRMWVT 139


RESULT   2
Q99736
ID   Q99736      PRELIMINARY;      PRT;  1928 AA.
AC   Q99736;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HSGCN1 (FRAGMENT).
GN   HSGCN1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SKELETAL MUSCLE;
RX   MEDLINE=97378024; PubMed=9234705;
RA   Marton M.J., Vazquez de Aldana C.R., Qiu H., Chakraburtty K.,
RA   Hinnebusch A.G.;
RT   "Evidence that GCN1 and GCN20, translational regulators of GCN4,
RT   function on elongating ribosomes in activation of eIF2alpha kinase
RT   GCN2.";
RL   Mol. Cell. Biol. 17:4474-4489(1997)
DR   EMBL; U77700; AAC51648.1; -.
DR   InterPro; IPR000357; HEAT_repeat.
DR   InterPro; IPR000847; HTH_LysR.
DR   Pfam; PF02985; HEAT; 7
DR   PROSITE; PS50077; HEAT_REPEAT; 3.
DR   PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
FT   NON_TER       1      1
SQ   SEQUENCE   1928 AA;  211510 MW;  8B8207D0C7A8FEC8 CRC64;


  Query Match             83.7%;  Score 41;  DB 4;  Length 1928;
  Best Local Similarity   87.5%;  Pred. No. 22;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWV 8
           |||||:||
Db     284 NLLRRLWV 291


RESULT   3
O95001
ID   O95001      PRELIMINARY;      PRT;  2392 AA.
AC   O95001;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   WUGSC:H_267D11.3 PROTEIN (FRAGMENT).
GN   WUGSC:H_267D11.3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kohlling T., Clarke K., Bauer C., Morris M.;
RT   "The sequence of Homo sapiens PAC clone 267D11.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Waterston R.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004812; AAC83183.1; -.
DR   InterPro; IPR000357; HEAT_repeat.
DR   InterPro; IPR000847; HTH_LysR.
DR   InterPro; IPR000663; Natr_peptide.
DR   Pfam; PF02985; HEAT; 7.
DR   PRINTS; PR00710; NATPEPTIDES.
DR   PROSITE; PS50077; HEAT_REPEAT; 3.
DR   PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
FT   NON_TER       1      1
SQ   SEQUENCE   2392 AA;  262292 MW;  1B15A227CCA33BA9 CRC64;


  Query Match             83.7%;  Score 41;  DB 4;  Length 2392;
  Best Local Similarity   87.5%;  Pred. No. 27;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWV 8
           |||||:||
Db     846 NLLRRLWV 853


RESULT   4
Q92616
ID   Q92616      PRELIMINARY;      PRT;  2412 AA.
AC   Q92616;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA0219 PROTEIN (FRAGMENT).
GN   KIAA0219.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RX   MEDLINE=97191544; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Ohara O.,
RA   Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   the coding sequences of 80 new genes (KIAA0201 KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).
DR   EMBL; D86973; BAA13209.1; -.
DR   InterPro; IPR000357; HEAT_repeat.
DR   InterPro; IPR000847; HTH_LysR.
DR   InterPro; IPR000663; Natr_peptide.
DR   Pfam; PF02985; HEAT; 7.
DR   PRINTS; PR00710; NATPEPTIDES.
DR   PROSITE; PS50077; HEAT_REPEAT; 3.
DR   PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
FT   NON_TER       1      1
SQ   SEQUENCE   2412 AA;  264339 MW;  6641FCD6DD8EE09DB CRC64;


  Query Match             83.7%;  Score 41;  DB 4;  Length 2412;
  Best Local Similarity   87.5%;  Pred. No. 27;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWV 8
           |||||:||
Db     866 NLLRRLWV 873


RESULT   5
Q9P3C0
ID   Q9P3C0      PRELIMINARY;      PRT;   390 AA.
AC   Q9P3C0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   B7N4.40.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
```

```
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL390218; CAB99235.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   390 AA;  44439 MW;  E1BE513D57672BD9 CRC64;


  Query Match             73.5%;  Score 36;  DB 3;  Length 390;
  Best Local Similarity   85.7%;  Pred. No. 42;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0,  Gaps    0;

Qy       1 NLLRRMW 7
           | |||||
Db     272 NFLRRMW 278


RESULT   6
Q9SND0
ID   Q9SND0      PRELIMINARY;      PRT;  3071 AA.
AC   Q9SND0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 343.7 KDA PROTEIN.
GN   F11C1_220.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Barques M., Collado M.C., Navarro P., Torel J., Perez-Alonso M.,
RA   Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X., Quetier F.,
RA   Salanoubat M.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL132976; CAB62317.1; -
DR   InterPro; IPR000005; HTHAraC.
DP   PPROSITE; PS00041; HTH_ARAC_FAMILY_1; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   3071 AA;  343688 MW;  4A3F7C9F71C92A25 CRC64;


  Query Match             73.5%;  Score 36;  DB 10;  Length 3071;
  Best Local Similarity   75.0%;  Pred. No. 3.2e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0,  Gaps    0;

Qy       1 NLLRRMWV 8
           |||| ||:
Db     601 NLLREMWL 608


RESULT   7
Q9LG50
ID   Q9LG50      PRELIMINARY;      PRT;  1292 AA.
AC   Q9LG50;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ESTS AU078742(C11888)
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0599D11.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002817; BAB03441.1; -.
DR   InterPro; IPR000767; Disease_resist.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003591; LRR_out.
DR   InterPro; IPR002182; NB-ARC.
DR   Pfam; PF00560; LRR; 14.
DR   Pfam; PF00931; NB-ARC; 1.
DR   PRINTS; PR00364; DISEASERSIST.
DR   SMART; SM00370; LRR; 5.
SQ   SEQUENCE   1292 AA;  144595 MW;  3DBAB30DF947F767 CRC64;


  Query Match             71.4%;  Score 35;  DB 10;  Length 1292;
  Best Local Similarity   62.5%;  Pred. No. 2.2e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWV 8
           :|||:|:
Db    1158 SLLRRLWI 1165


RESULT   8
Q25913
ID   Q25913      PRELIMINARY;      PRT;  332 AA.
AC   Q25913;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MALARIA ANTIGEN (FRAGMENT).
GN   M26.32.12.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MALAYAN CAMP;
RA   MEDLINE 91215052, PubMed 1726149;
RA   Cheng Q., Jones G., Liu E.X., Kidson C., Saul A.;
RT   "Identification of a common plasmodium falciparumepitope (CPE)
RT   recognized by a pan specific inhibitorymonoclonal antibody.";
RL   Mol. Biochem. Parasitol. 49:73-82(1991).
DR   EMBL; M63277; AAA29642.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   332 AA;  37545 MW;  7229B59180C897A8 CRC64;


  Query Match             69.4%;  Score 34;  DB 5;  Length 332;
  Best Local Similarity   71.4%;  Pred. No. 62;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMW 7
           |||::||
Db     124 NLLKKMW 130


RESULT   9
Q96L58
ID   Q96L58      PRELIMINARY;      PRT;  329 AA.
AC   Q96L58;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
```

```
DE   BETA-1,3-GALACTOSYLTRANSFERASE-6.
GN   B3GALT6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bai X., Zhou D., Brown J.R., Hennet T., Esko J.D.;
RT   "Biosynthesis of the linkage region of glycosaminoglycans: Cloning and
RT   activity of galactosyltransferase II, the sixth member of the beta-
RT   1,3-galactosyltransferase family (beta3GalT6).";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY050570; AAL11442.1; -.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   329 AA;  37123 MW;  3753F61B00DEA313 CRC64;


  Query Match             69.4%;  Score 34;  DB 4;  Length 329;
  Best Local Similarity   85.7%;  Pred. No. 87;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMW 7
           ||||| |
Db       2 NLLRRAW 8


RESULT  10
O02310
ID   O02310      PRELIMINARY;      PRT;   338 AA.
AC   O02310;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   T09E11.10 PROTEIN.
GN   T09E11.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   McLay K.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z81147; CAB03537.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
SQ   SEQUENCE   338 AA;  38916 MW;  2471F568F402436D CRC64;


  Query Match             69.4%;  Score 34;  DB 5;  Length 338;
  Best Local Similarity   55.6%;  Pred. No. 90;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           |:||: |:|
Db      96 NVLRKTWMT 104


RESULT  11
Q93CU8
ID   Q93CU8      PRELIMINARY;      PRT;   409 AA.
AC   Q93CU8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   O-ANTIGEN FLIPPASE.
GN   WZX.
OS   Shigella boydii.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Shigella.
OX   NCBI_TaxID=621;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21481970; PubMed=11598067;
RA   Wang L., Qu W., Reeves P.R.;
RT   "Sequence Analysis of Four Shigella boydii O Antigen Loci: Implication
RT   for Escherichia coli and Shigella Relationships.";
RL   Infect. Immun. 69:6423-6430(2001).
DR   EMBL; AF402312; AAL27316.1; -.
SQ   SEQUENCE   409 AA;  47677 MW;  875B820EAC5CB31A CRC64;


  Query Match             69.4%;  Score 34;  DB 2;  Length 409;
  Best Local Similarity   62.5%;  Pred. No. 1.1e-02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWV 8
           |||| :|:
Db     198 NLLRNVWI 205


RESULT  12
O46945
ID   O46945      PRELIMINARY;      PRT;   645 AA.
AC   O46945;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NADH DEHYDROGENASE SUBUNIT F (FRAGMENT).
GN   NDHF.
OS   Heliotropium arborescens.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; euasterids I incertae sedis; Boraginaceae;
OC   Heliotropium.
OX   NCBI_TaxID=28508;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ferguson D.M.;
RT   "Phylogenetic analysis and circumscription of Hydrophyllaceae based on
RT   ndhF sequence data.";
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Schranz E., Olmstead R.G.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.
DR   EMBL; AF014000; AAB94517.2; -.
DR   InterPro; IPR001750; Oxidored_q1.
DR   InterPro; IPR002128; Oxidored_q1_C.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF01010; oxidored_q1_C; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
KW   Chloroplast; NAD; Oxidoreductase; Plastoquinone.
FT   NON_TER       1       1
FT   NON_TER     645     645
SQ   SEQUENCE   645 AA;  72688 MW;  1CD1415B78EAD194 CRC64;


  Query Match             69.4%;  Score 34;  DB 8;  Length 645;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 LRRMWV 8
           ||||||
Db      28 LRRMWV 33
```

```
RESULT 13
Q97RR4
ID   Q97RR4      PRELIMINARY;      PRT;   660 AA.
AC   Q97RR4;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CATION-TRANSPORTING ATPASE, E1-E2 FAMILY.
GN   SP0729.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TIGR4;
RX   MEDLINE=21357209; PubMed=11463916;
RA   Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA   Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA   Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA   Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA   Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA   McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA   Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA   Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT   "Complete genome sequence of a virulent isolate of Streptococcus
RT   pneumoniae.";
RL   Science 293:498-506(2001).
DR   EMBL; AE007381; AAK74870.1; -.
DR   TIGR; SP0729; -.
DR   InterPro; IPR001756; Cu_ATPase.
DR   InterPro; IPR001757; E1-E2_ATPase.
DR   InterPro; IPR001454; Hydrolase.
DR   InterPro; IPR000150; Hypothet_cof.
DR   Pfam; PF00122; E1-E2_ATPase; 1.
DR   Pfam; PF00702; Hydrolase; 1.
DR   PRINTS; PR00119; CATATPASE.
DR   PRINTS; PR00943; CUATPASE.
DR   PROSITE; PS00154; ATPASE_E1_E2; UNKNOWN_1.
DR   PROSITE; PS01229; COF_2; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   660 AA;  70670 MW;  0A98266A9B470AD CRC64;


  Query Match             69.4%;  Score 34;  DB 16;  Length 660;
  Best Local Similarity   55.6%;  Pred. No. 1.7e+02;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           :||:| |:|
Db      19 SLLQRFWIT 27


RESULT 14
Q9TIT9
ID   Q9TIT9      PRELIMINARY;      PRT;   700 AA.
AC   Q9TIT9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NADH DEHYDROGENASE (FRAGMENT).
GN   NDHF.
OS   Tournefortia acutiflora.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Hydrophyllaceae; Tournefortia.
OX   NCBI_TaxID=79404;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ferguson D.M.;
RT   "Phylogenetic analysis and circumscription of Hydrophyllaceae based on
RT   ndhF sequence data.";
RL   Syst. Bot. 0:0-0(1998).
CC   -!- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.
DR   EMBL; AF047813; AAF22420.1; -.
DR   InterPro; IPR001750; Oxidored_q1.
DR   InterPro; IPR022128; Oxidored_q1_C.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF01010; oxidored_q1_C; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
KW   Chloroplast; NAD; Oxidoreductase; Plastoquinone.
FT   NON_TER       1       1
FT   NON_TER     700     700
SQ   SEQUENCE   700 AA;  78889 MW;  96BA728EAEA7570C CRC64;


  Query Match             69.4%;  Score 34;  DB 8;  Length 700;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 LPRMWV 9
           ||||||
Db      28 LPRMWV 33


RESULT 15
Q9TL48
ID   Q9TL48      PRELIMINARY;      PRT;   706 AA.
AC   Q9TL48;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NADH DEHYDROGENASE SUBUNIT F (FRAGMENT).
GN   NDHF.
OS   Rhamnus davurica.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Rosales; Rhamnaceae; Rhamnus.
OX   NCBI_TaxID=105902;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20448987; PubMed=10877943;
RA   Olmstead R.G., Kim K.J., Jansen R.K., Wagstaff S.J.;
RT   "The phylogeny of the asteridae sensu lato based on chloroplast ndhF
RT   gene sequences.";
RL   Mol. Phylogenet. Evol. 16:96-112(2000).
CC   -!- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.
DR   EMBL; AF130225; AAF08181.1; -.
DR   InterPro; IPR001750; Oxidored_q1.
DR   InterPro; IPR022128; Oxidored_q1_C.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF01010; oxidored_q1_C; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
KW   Chloroplast; NAD; Oxidoreductase; Plastoquinone.
FT   NON_TER       1       1
FT   NON_TER     706     706
SQ   SEQUENCE   706 AA;  79608 MW;  B96A7EA6448EC9AE CRC64;


  Query Match             69.4%;  Score 34;  DB 9;  Length 706;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 LRRMWV 8
           ||||||
Db      27 LRRMWV 32



Search completed: September 5, 2002, 15:31:01
```

Job time: 466 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:26:08 ; Search Time 158.43 Seconds
(without alignments)
6.310 Million cell updates/sec

Title: US-09-744-804-36
Perfect score: 49
Sequence: 1 NLLRPMWVT 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	49	100.0	9	21	AAY82841	Lactadherin (BA-46
2	49	100.0	387	16	AAR77252	HMFG 46 kDa antige
3	49	100.0	487	21	AAY94453	Human lactadherin
4	41	83.7	1763	22	AAM23507	Human EST encoded
5	41	83.7	2473	22	AAM78911	Human protein SEQ
6	41	83.7	2668	22	AAM79895	Human protein SEQ
7	36	73.5	112	22	AAU23593	Novel human enzyme
8	35	71.4	666	22	ABG28776	Novel human diagno
9	34	69.4	329	21	AAY51901	Human 3GnT protein
10	34	69.4	750	22	AAU37812	Streptococcus pneu
11	34	69.4	798	22	ABG19088	Novel human diagno
12	33	67.3	187	22	AAG65919	Amino acid sequenc
13	33	67.3	204	21	AAC28778	Arabidopsis thalia
14	33	67.3	234	22	AAB90765	Human shear stress
15	33	67.3	258	21	AAC28777	Arabidopsis thalia
16	33	67.3	293	22	AAG91417	C glutamicum prote
17	33	67.3	293	22	AAB78933	C. glutamicum SRT
18	33	67.3	304	21	AAG28776	Arabidopsis thalia
19	33	67.3	1076	22	ABB59684	Drosophila melanog
20	33	67.3	4010	22	ABB61520	Drosophila melanog
21	32	65.3	22	20	AAY19350	Lecithin:cholester
22	32	65.3	22	20	AAY19096	Lecithin:cholester
23	32	65.3	22	20	AAY18842	Lecithin:cholester
24	32	65.3	22	20	AAY18579	Lecithin:cholester
25	32	65.3	29	22	ABG29861	Novel human diagno
26	32	65.3	57	21	AAC52749	Arabidopsis thalia
27	32	65.3	72	21	AAG52747	Arabidopsis thalia
28	32	65.3	94	22	AAU66974	Propionibacterium
29	32	65.3	117	22	AAU40249	Propionibacterium
30	32	65.3	160	22	AAY65412	Propionibacterium
31	32	65.3	273	22	AAB79239	Corynebacterium gl
32	32	65.3	308	22	AAU19724	Human novel extrac
33	32	65.3	321	21	AAG46935	Arabidopsis thalia
34	32	65.3	356	22	ABG13349	Novel human diagno
35	32	65.3	371	22	AAG91558	C glutamicum prote
36	32	65.3	376	22	ABB63737	Drosophila melanog
37	32	65.3	390	22	ABG05013	Novel human diagno
38	32	65.3	424	22	ABG13423	Novel human diagno
39	32	65.3	440	21	AAG46934	Arabidopsis thalia
40	32	65.3	456	21	AAG46943	Arabidopsis thalia
41	32	65.3	600	20	AAY34979	Protein which is s
42	32	65.3	720	20	AAW94905	Mouse pheromone re
43	32	65.3	864	20	AAW94914	Mouse pheromone re
44	32	65.3	852	20	AAW94904	Mouse pheromone re
45	32	65.3	946	18	AAW27985	Human cytomegalovi

ALIGNMENTS

RESULT 1
AAY82841
ID AAY82841 standard; peptide; 9 AA.
XX
AC AAY82841;
XX
DT 14-JUN-2000 (first entry)
XX
DE Lactadherin (BA-46) peptide fragment (tumour associated antigen).
XX
KW tumour associated antigen peptide; BAA; cancer; carcinoma;
KW treatment; prevention; cure; anti-tumour vaccine; metastases;
KW breast; bladder; prostate; pancreas; ovary; thyroid; colon;
KW stomach; carcinoma; MHC Class I; HLA-A2; human;
KW Major Histocompatibility Complex; uteroglobin;
KW prostate specific antigen; prostate specific membrane antigen;
KW prostate acid phosphatase; mucin; lactadherin;
KW teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRIPTO-1.
XX
OS Homo sapiens.
XX
PN WO200006723-A1.
XX
PD 10-FEB-2000
XX
PF 29-JUL-1999; 99WO-IL00417.
XX
PR 30-JUL-1998; 98IL-0125608.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (BIOT-) BIO-TECHNOLOGY GEN CORP.
XX
PI Eisenbach L, Carmon L, Tirosh B, Bar haim E, Paz A, Fridkin M;
PI Fitzer-attas C;

```
XX
DR   WPI; 2000-205463/18.
XX
PT   Tumor associated antigen peptides, especially derived from uroplakin,
PT   useful as vaccines to prevent or cure cancers including breast,
PT   bladder, prostate, pancreas, ovary, thyroid, colon and stomach -
XX
PS   Claim 17; Page 99; 113pp; English.
XX
CC   Tumour associated antigen peptides (TAA) may be used for the
CC   treatment, prevention and cure of cancer or cancer metastases. The
CC   cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
CC   colon, stomach, head or neck cancer or a carcinoma. The tumour
CC   associated antigens are presentable to the immune system by HLA A2
CC   molecules and are generally between 8 to 10 amino acids in length.
CC   The amino acids located at positions 2 and 9 of the tumour associated
CC   antigens are the anchor residues which participate in the binding to
CC   MHC class I molecules, more specifically HLA-A2.  More tumour
CC   associated antigens are described in GENESEQ records AAY82805 Y82882.
CC   Those tumour associated antigens described in records AAY82806-Y82824
CC   and AAY82855-YY82869 are derived from Uroplakin, such as Uroplakin II,
CC   Uroplakin Ia, Uroplakin III and Uroplakin Ib.  Those described in
CC   records AAY82825-Y82829 are derived from prostate specific antigen
CC   (PSA). Those described in records AAY82830-Y82835 are derived from
CC   prostate specific membrane antigen (PSMA).  Those described in
CC   records YY82836 AAY82839 are derived from prostate acid phosphotase
CC   (PAP)  Those described in records AAY82840-Y82846 are derived
CC   from Lactadherin (BA-46)  Those described in records AAY82847-Y82854
CC   are derived from Mucin and those described in records AAY82871-Y82882
CC   are derived from Teratocarcinoma derived growth factor (CRIPTO-1).
XX
SQ   Sequence   9 AA;


  Query Match             100.0%;  Score 49;  DB 21;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           |||||||||
Db       1 nllrrmwvt 9


RESULT   2
AAR77252
ID   AAR77252 standard; Protein; 387 AA
XX
AC   AAR77252;
XX
DT   21-NOV-1995  (first entry)
XX
DE   HMFG 46 kDa antigen.
XX
KW   HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
KW   epithelium; tumor; breast cancer; monoclonal antibody; MAb.
XX
OS   Homo sapiens.
XX
PN   WO9515171-A.
XX
PD   08-JUN-1995.
XX
PF   05-DEC-1994;   94WO-US13967.
XX
PR   03-DEC-1993;   93US-0162402.
XX
PA   (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI   Ceriani RL,  Larocca DJ,  Peterson JA;
XX
DR   WPI; 1995-215151/28.
DR   N-PSDB; AAQ91198.
XX
PT   46 kD apparent molecular weight human milk fat globule antigen -
PT   used in assays to determine the presence of a cancerous tumour of
PT   epithelial origin, and in a vaccine against neoplastic tumours
XX
PS   Claim 6; Page 46-47; 68pp; English.
XX
CC   A complete cDNA sequence for the 46 kDa HMFG antigen, a major
CC   component of the apical surface of the normal breast epithelial
CC   cell, was obtained by PCR and RACE methods.  cDNA clones can be
CC   used to prepare MAbs for use in immunotherapy, immunohistopathology,
CC   prognosis, diagnosis, imaging and therapy.  Recombinant antigen can
CC   be expressed in prokaryotic or (glycosylated) in eukaryotic cells.
XX
SQ   Sequence   387 AA;


  Query Match             100.0%;  Score 49;  DB 16;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.42;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWVT 9
           |||||||||
Db     131 nllrrmwvt 139


RESULT   3
AAY94453
ID   AAY94453 standard; Protein; 387 AA
XX
AC   AAY94453;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Human lactadherin protein.
XX
KW   Human; lactadherin; MGF-E8; anti-tumour; immune response;
KW   exosome; dendritic cell.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Secretion_signal
FT   Protein         24..387
FT                   /label= Lactadherin
FT   Binding-site    46..48
FT                   /label= Integrin_binding_site
XX
PN   EP1004664-A1.
XX
PD   31-MAY-2000.
XX
PF   24-NOV-1998;   98EP-0402925.
XX
PR   24-NOV-1998;   98EP-0402925.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (CURI-) INST CURIE.
XX
DR   WPI; 2000-352597/31.
DR   N-PSDB; AAA27140.
XX
PT   Chimeric isolated (human) lactadherin polypeptide that functions as an
PT   adaptor of cross-priming to eliminate pathogenic antigens, e.g. in
PT   tumors -
XX
PS   Example 3; Page 12; 20pp; English.
XX
CC   Lactadherin protein was found in exosomes produced by dendritic cells.
CC   The protein is involved in the phagocytosis of particulate antigens by
CC   dendritic cells. Exosomes produced by dendritic cells exposed to
CC   tumour antigens induce potent immune responses. Lactadherin or variants
CC   of it may be used in the mediation of an immune response. Variants of
```

CC lactadherin may be used for inhibition and/or stimulation of the
CC cross-priming of antigens and stimulation of the phagocytosis of
CC antigens by dendritic cells. Compositions derived from lactadherin can
CC also be used to monitor an immune response, more specifically a CTL
CC (cytotoxic T-lymphocyte) response and also to produce CTLs specific
CC for a selected antigen. The present sequence is the human
CC lactadherin protein.
XX
SQ Sequence 387 AA;

Query Match 100.0%; Score 49; DB 21; Length 387;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLLRRMWVT 9
|||||||||
Db 131 nllrrmwvt 139

RESULT 4
AAM23507
ID AAM23507 standard; Protein; 1763 AA.
XX
AC AAM23507;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1032.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse,
KW tomato; monkey, dog, sea urchin, expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX
OS Homo sapiens
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746
PR 03-AUG-2000; 2000US-0631451
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen F, Asundi V,
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI; 2001-476164/51
DR N-PSDB; AAH98166.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use
XX
PS Claim 20; Page 788 792; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SQ Sequence 1763 AA;

Query Match 83.7%; Score 41; DB 22; Length 1763;
Best Local Similarity 87.5%; Pred. No. 53;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLLRRMWV 8
|||||:||
Db 217 nllrrlwv 224

RESULT 5
AAM78911
ID AAM78911 standard; Protein; 2473 AA.
XX
AC AAM78911;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1573.
XX
KW Human; cytokine, cell proliferation, cell differentiation; gene therapy;
KW vaccine, peptide therapy, stem cell growth factor, haematopoiesis;
KW tissue growth factor, immunomodulatory, cancer, leukaemia;
KW nervous system disorder, arthritis, inflammation.
XX
OS Homo sapiens.
XX
PN WO200175190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang B, Wang J, Zhang J, Fan P, Chen F, Wang ZW;
PI Xue AJ, Yang Y, Wojhrman T, Goodrich P;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK52044.
XX
PT Nucleic acids encoding polypeptides with cytokine like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 3931 3935; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78322-AAM80792) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity, and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80920) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 2473 AA;

Query Match 83.7%; Score 41; DB 22; Length 2473;
Best Local Similarity 87.5%; Pred. No. 75;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLLRRMWV 8
||||||:|
Db 927 nllrrlwv 934

RESULT 6
AAM79895
ID AAM79895 standard; Protein; 2668 AA.
XX
AC AAM79895;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3541.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy, stem cell growth factor; haematopoiesis;
KW tissue growth factor, immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663571.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE) HYSEQ INC.
XX
PI Tang YT, Liu C, Ormanac PT, Asundi V, Zhou P, Xu C, Cao Y, Mu Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen F, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich P;
XX
DR WPI; 2001-476283/51
DR N-PSDB; AAK53028
XX
PT Nucleic acids encoding polypeptides with cytokine like activities
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 377-379; 622lpp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323 AAM80020) that exhibit activity relating to
CC cytokines, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3555
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 2668 AA;

Query Match 83.7%; Score 41; DB 22; Length 2668;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLLRRMWV 8
||||||:|
Db 1122 nllrrlwv 1129

RESULT 7
AAU23593
ID AAU23593 standard; Protein; 112 AA.
XX
AC AAU23593;
XX
DT 18 DEC 2001 (first entry)
XX
DE Novel human enzyme polypeptide #879.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190326.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 25-JUL-2000; 2000US-0220963.
PR 25-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229245.
PR 05-SEP-2000; 2000US-0229609.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230447.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231342.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232407.
PR 14-SEP-2000; 2000US-0232408.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236832.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246614.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI, 2001-455566/50.
DR N-PSDB; AAS41469.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -
XX
PS Claim 11; SEQ ID No 1589; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785 AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAB22815 AAB22814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 112 AA,

Query Match 73.5%, Score 26, DB 22, Length 112;
Best Local Similarity 75.0%, Pred. No. 28;
Matches 6, Conservative 1, Mismatches 1, Indels 0, Gaps 0;

Qy 1 NILERMWV 8
||| |:||
Db 5 NLLERLWV 12

RESULT 8
ABG28776

```
ID   ABG28776 standard; Protein; 666 AA.
XX
AC   ABG28776;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #28767.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement, medical imaging, diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-APR-2000; 2000US-0640167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS62963.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID NO 59135; 104pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   666 AA;


  Query Match             71.4%;  Score 35;  DB 22;  Length 666;
  Best Local Similarity   75.0%;  Pred. No. 2.5e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMWV 8
            |:|| ||:
Db     137 nllrimwi 144


RESULT   9
AAY51901
ID   AAY51901 standard; Protein; 329 AA.
XX
AC   AAY51901;
XX
DT   15-JUN-2000  (first entry)
XX
DE   Human 3GnT protein.
XX
KW   UDP (uridine diphosphate)-N-acetylglucosaminyl; beta1,4 galactoside;
KW   beta1,3-N acetylglucosaminyl transferase; 3GnT; polylactosaminyl;
KW   disaccharide production; oligosaccharide production; glycan;
KW   glycoprotein; gene therapy; transgenic animal; human.
XX
OS   Homo sapiens.
XX
PN   WO200011190-A2.
XX
PD   02-MAR-2000.
XX
PF   06-AUG-1999;   99WO-CH00365.
XX
PR   28-AUG-1998;   98CH-0001717.
XX
PA   (BERG/) BERGER E G.
PA   (HENN/) HENNET T.
XX
PI   Berger EG,  Hennet T;
XX
DR   WPI; 2000-224707/19.
DR   N-PSDB; AAZ89455.
XX
PT   New nucleic acid encoding a glucosaminyl transferase, used for in vivo
PT   or in vitro synthesis of glycan chains containing a specific
PT   disaccharide repeat element -
XX
PS   Disclosure; Page 31-32; 34pp; German.
XX
CC   This invention describes a novel isolated nucleic acid (I) encoding
CC   UDP (uridine diphosphate) N acetylglucosaminyl:beta1,4
CC   galactosidebeta1,3-N-acetylglucosaminyl transferase (3GnT)
CC   (polylactosaminyl type). 3GnTs are used, in vivo or in vitro, for
CC   enzymatic production of di- or oligo-saccharides, particularly glycans,
CC   containing the repeated disaccharide motif GlcNAcbeta1-3Gal. (I) is
CC   useful for recombinant expression of 3GnTs, to alter adhesion and
CC   recognition properties of transfected cells and to produce glycoproteins
CC   substituted by polylactosaminoglycans, when introduced into cells that
CC   express glycoproteins. Vectors containing (I) are used in gene therapy
CC   and for production of transgenic animals. Fragments of (I) are used to
CC   detect related sequences from other species and as antisense reagents.
CC   This sequence represents the human 3GnT protein described in the
CC   invention.
XX
SQ   Sequence   329 AA;


  Query Match             69.4%;  Score 34;  DB 21;  Length 329;
  Best Local Similarity   85.7%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLLRRMW 7
            ||||| |
Db       2 nllrraw 8


RESULT  10
AAU37812
ID   AAU37812 standard; Protein; 750 AA.
XX
AC   AAU37812;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Streptococcus pneumoniae cellular proliferation protein #241.
XX
```

KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Streptococcus pneumoniae.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206948P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS55671.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3, Seq ID No 13405, 511pp, English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 750 AA;

Query Match 69.4%; Score 34; DB 22; Length 750;
Best Local Similarity 55.6%; Pred. No. 4.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLLRRMWVT 9
:|||:| |:|
Db 109 sllqrfwit 117

RESULT 11
ABG19088
ID ABG19088 standard; Protein; 798 AA.
XX
AC ABG19088;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #19079.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-649462/73.
DR N-PSDB; AAS83275.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 49447; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 798 AA;

Query Match 69.4%; Score 34; DB 22; Length 798;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLRRMW 7
||||||
Db 362 llrrmw 367

RESULT 12
AAG65919
ID AAG65919 standard; Protein; 187 AA.
XX
AC AAG65919;
XX
DT 11-FEB-2002 (first entry)
XX
DE Amino acid sequence of novel human arginine-rich protein (ARP).
XX

```
KW   ARP; angiogenesis; vascular endothelial growth factor; VEGF; cytostatic;
KW   arginine-rich protein; cardiant; antirheumatic; antiarthritic; human;
KW   antiatherosclerotic; vasotropic; gynecological; antidiabetic; vulnerary;
KW   antiulcer; dermatological; ophthalmological; antipsoriatic; apoptosis;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 16
FT                   /note  "encoded by CGC"
FT   Misc-difference 17
FT                   /note= "encoded by GTG"
FT   Misc-difference 48
FT                   /note= "encoded by AAA"
FT   Misc-difference 49
FT                   /note= "encoded by GAC"
FT   Misc-difference 50..51
FT                   /note= "nucleotides encoding these 2 residues are not
FT                          indicated in the corresponding nucleotide
FT                          sequence"
XX
PN   WO200170174-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09043.
XX
PR   21-MAR-2000; 2000US-191201P
XX
PA   (CURA-) CURAGEN CORP.
PA   (GETH ) GENENTECH INC
XX
PI   Rastelli L.K.; Gerber H;
XX
DR   WPI; 2001-639087/73.
DR   N-PSDB; AA167447.
XX
PT   Modulating angiogenesis and/or apoptosis for preventing or treating
PT   cancer, myocardial infarction and promoting healing, by modulating the
PT   activity of vascular endothelial growth fit related gene
PT   polypeptide -
XX
PS   Claim 4; Page 15; 155pp; English.
XX
CC   The invention relates to modulating angiogenesis and cell survival that
CC   involves modulating the activity of at least one vascular endothelial
CC   growth factor (VEGF)-modulated gene polypeptide. The method is useful for
CC   modulating angiogenesis and cell survival, for treating tumour and cancer
CC   by decreasing angiogenesis in cancerous tumours and treating myocardial
CC   infarction and promoting healing, by increasing angiogenesis. Transgenic
CC   non-human animals, having disrupted arginine-rich protein (ARP), are
CC   useful for determining the clinical stage of ovarian tumorous, which is
CC   useful for determining it the tumour has potential for metastasis. ARP is
CC   useful in gene therapy and in diagnostic applications. VEGFmq proteins
CC   are useful in the treatment of tumours, neoplasias, hemangiomas,
CC   rheumatoid arthritis, atherosclerosis, idiopathic pulmonary fibrosis,
CC   vascular restenosis, arteriovenous malformations, meningioma, neovascular
CC   glaucoma, psoriasis, hemophilic joints, hypertrophic scars, Osler-Weber
CC   syndrome, scleroderma, vascular adhesion pathologies, synovitis,
CC   dermatitis, endometriosis, diabetic retinopathy, neovascularization
CC   associated with corneal injury or grafts, wound, sore, and ulcer healing.
CC   The present sequence represents the amino acid sequence of human ARP.
XX
SQ   Sequence   187 AA;

  Query Match             67.3%;  Score 33;  DB 22;  Length 187;
  Best Local Similarity   71.4%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 LRRMWVT 9
           :|||| |
Db       4 MRRMWAT 10


RESULT 13
AAG28778
ID   AAG28778 standard; Protein; 204 AA.
XX
AC   AAG28778;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 34124.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
```

PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140941.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142265.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143077.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151940.
PR 07-SEP-1999; 99US-0152463.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0154758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158869.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161459.
PR 26-OCT-1999; 99US-0161460.
PR 26-OCT-1999; 99US-0161461.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 67.4%; Score 33; DB 21; Length 204;
Best Local Similarity 62.5%; Pred. No. 1.8e-02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NILRRMWV 8
|: ||:||
Db 22 nvtrriwv 29

RESULT 14
AAB90765
ID AAB90765 standard; Protein; 234 AA.
XX

```
AC   AAB90765;
XX
DT   15-JUN-2001  (first entry)
XX
DE   Human shear stress-response protein SEQ ID NO: 30.
XX
KW   Human; shear stress-response protein; vascular disease;
KW   arteriosclerosis.
XX
OS   Homo sapiens.
XX
PN   WO200125427-A1
XX
PD   12-APR-2001.
XX
PF   02-OCT-2000; 2000WO-JP06840
XX
PR   01-OCT-1999;   99JP-0280976.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (NOJI/) NOJIMA H.
XX
PI   Nojima H,  Yoshisue H,  Obayashi M,  Ota T,  Kawabata A,  Sakurada F;
PI   Koga T,  Sekine S,  Nakamura Y,  Sugano S.
XX
DR   WPI; 2001-266308/27.
DR   N-PSDB; AAH02888.
XX
PT   DNA sequences, proteins encoded by them and antibodies against them
PT   useful in diagnosis and treatment of vascular disease caused by
PT   arteriosclerosis -
XX
PS   Claim 60; Page 258-259; 678pp; Japanese.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human shear stress response proteins. These are useful in the
CC   diagnosis, treatment and screening of vascular diseases caused by
CC   arteriosclerosis, including heart failure, post-PTCA restenosis and
CC   hypertension.
XX
SQ   Sequence   234 AA;


  Query Match          67.3%;  Score 33;  DB 22;  Length 234;
  Best Local Similarity   71.4%;  Pred. No. 2e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 LKRMWVT 9
           :|||| |
Db      53 mrmwat 59


RESULT  15
AAG28777
ID   AAG28777 standard; Protein; 258 AA.
XX
AC   AAG28777;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 34123
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301489
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US 0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US 0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132864.
PR   11-MAY-1999;   99US 0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14 MAY 1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US 0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
```

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 67.3%; Score 33; DB 21; Length 258;
Best Local Similarity 62.5%; Pred. No. 2.4e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLLRRMWV 8
|: ||:||
Db 76 nvtrriwv 83

Search completed: September 5, 2002, 15:26:29
Job time: 369 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:27:34 ; Search time 58.86 Seconds
(without alignments)
3.735 Million cell updates/sec

Title: US-09-744-804-37
Perfect score: 45
Sequence: 1 NLPETPILA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	100.0	159	2	US-08-162-402B-12	Sequence 12, Appl
2	45	100.0	217	1	US-07-607-538C-3	Sequence 3, Appli
3	45	100.0	217	2	US-08-162-402B-3	Sequence 3, Appli
4	45	100.0	218	1	US-07-607-538C-2	Sequence 2, Appli
5	45	100.0	218	2	US-08-162-402B-2	Sequence 2, Appli
6	45	100.0	387	2	US-08-162-402B-6	Sequence 6, Appli
7	45	100.0	465	2	US-08-162-402B-8	Sequence 8, Appli
8	37	82.2	160	2	US-08-162-402B-10	Sequence 10, Appl
9	32	71.1	14	2	US-08-162-402B-26	Sequence 26, Appl
10	32	71.1	590	2	US-08-756-317-12	Sequence 12, Appl
11	32	71.1	3672	2	US-08-822-445-12	Sequence 12, Appl
12	32	71.1	3672	4	US-09-396-540-12	Sequence 12, Appl
13	32	71.1	3801	2	US-08-822-445-10	Sequence 10, Appl
14	32	71.1	3801	4	US-09-396-540-10	Sequence 10, Appl
15	31	68.9	85	2	US-08-480-229C-2	Sequence 2, Appli
16	31	68.9	85	2	US-08-659-235C-2	Sequence 2, Appli
17	31	68.9	109	1	US-08-111-939-25	Sequence 25, Appl
18	31	68.9	157	2	US-08-162-402B-13	Sequence 13, Appl
19	31	68.9	289	4	US-09-077-675A-5	Sequence 5, Appli
20	31	68.9	320	2	US-08-480-229C-20	Sequence 20, Appl
21	31	68.9	320	2	US-08-659-235C-20	Sequence 20, Appl
22	31	68.9	353	4	US-09-077-675A-3	Sequence 3, Appli
23	31	68.9	391	4	US-08-984-618-11	Sequence 11, Appl
24	31	68.9	464	2	US-08-162-402B-9	Sequence 9, Appli
25	30	66.7	109	1	US-08-111-939-20	Sequence 20, Appl
26	30	66.7	154	2	US-08-162-402B-17	Sequence 17, Appl
27	30	66.7	218	1	US-07-607-538C-5	Sequence 5, Appli
28	30	66.7	218	2	US-08-162-402B-5	Sequence 5, Appli
29	30	66.7	449	2	US-08-927-744-2	Sequence 2, Appli
30	30	66.7	1438	4	US-09-209-916-1	Sequence 1, Appli
31	30	66.7	1471	1	US-08-683-839B-3	Sequence 4, Appli
32	30	66.7	1661	2	US-08-882-083-2	Sequence 2, Appli
33	30	66.7	1661	2	US-08-558-107-2	Sequence 2, Appli
34	30	66.7	1661	4	US-09-243-539-2	Sequence 2, Appli
35	30	66.7	2332	1	US-07-864-004B-4	Sequence 4, Appli
36	30	66.7	2332	1	US-08-251-937A-4	Sequence 4, Appli
37	30	66.7	2332	1	US-08-212-139A-2	Sequence 2, Appli
38	30	66.7	2332	1	US-08-276-594A-2	Sequence 2, Appli
39	30	66.7	2332	1	US-08-474-503-2	Sequence 2, Appli
40	30	66.7	2332	2	US-08-670-707A-2	Sequence 2, Appli
41	30	66.7	2332	4	US-09-037-601-2	Sequence 2, Appli
42	30	66.7	2332	4	US-09-324-867-4	Sequence 4, Appli
43	30	66.7	2332	5	PCT-US93-04275-4	Sequence 4, Appli
44	30	66.7	2332	5	PCT-US94-12000-2	Sequence 2, Appli
45	30	66.7	2343	4	US-09-324-867-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-162-402B-12
; Sequence 12, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LARocca, DAVID J.
; TITLE OF INVENTION: Gr EGATION HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: PBG 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213 622-7700
; TELEFAX: 213 489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-12

Query Match 100.0%; Score 45; DB 2; Length 159;
Best Local Similarity 100.0%; Pred No. 0.038;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 NLFETPILA 9
          |||||||||
Db    128 NLFETPILA 136
```

```
RESULT  2
US-07-607-538C-3
; Sequence 3, Application US/07607538C
; Patent No. 5455031
;  GENERAL INFORMATION:
;    APPLICANT:  Ceriani Dr., Roberto L.
;    APPLICANT:  Peterson Dr., Jerry A.
;    APPLICANT:  Larocca, David J.
;    TITLE OF INVENTION:  POLYPEPTIDE WITH 46
;    TITLE OF INVENTION:  DIFFERENTIATION ANTIGEN BINDING SPECIFITY AND  CLOTTING
;    TITLE OF INVENTION:  FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
;    TITLE OF INVENTION:  FUSION  PROTEIN,  POLYNUCLEOTIDE AND  POLYRIBO-
;    TITLE OF INVENTION:  NUCLEOTIDE  ENCODING  THE POLYPEPTIDE, ANTI-
;    TITLE OF INVENTION:  POLYPEPTIDE ANTIBODIES, KITS  AND  METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  V.  Amzel & Assoc.
;      STREET:  2055 No. 54550311h Broadway
;      CITY:  Walnut Creek
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94596
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS 5.0
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/607,538C
;      FILING DATE:  01-NOV-1990
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Viviana Amzel
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  CPPCC-004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 943-1931
;      TELEFAX:  (510) 943-1189
;      TELEX:  N.A.
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  217 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:
US-07-607-538C-3
```

Query Match 100.0%; Score 45; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 NLFETPILA 9
          |||||||||
Db    186 NLFETPILA 194
```

```
RESULT  3
US-08-162-402B-3
; Sequence 3, Application US/08162402B
; Patent No. 5972347
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  217 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-3
```

Query Match 100.0%; Score 45; DB 2; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 NLFETPILA 9
          |||||||||
Db    186 NLFETPILA 194
```

```
RESULT  4
US-07-607-538C-2
; Sequence 2, Application US/07607538C
; Patent No. 5455031
;  GENERAL INFORMATION:
;    APPLICANT:  Ceriani Dr., Roberto L.
;    APPLICANT:  Peterson Dr., Jerry A.
;    APPLICANT:  Larocca, David J.
;    TITLE OF INVENTION:  POLYPEPTIDE WITH 46
;    TITLE OF INVENTION:  DIFFERENTIATION ANTIGEN BINDING SPECIFITY AND  CLOTTING
;    TITLE OF INVENTION:  FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
;    TITLE OF INVENTION:  FUSION  PROTEIN,  POLYNUCLEOTIDE AND  POLYRIBO-
;    TITLE OF INVENTION:  NUCLEOTIDE  ENCODING  THE POLYPEPTIDE, ANTI-
;    TITLE OF INVENTION:  POLYPEPTIDE ANTIBODIES, KITS  AND  METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  V.  Amzel & Assoc.
;      STREET:  2055 No. 54550311h Broadway
;      CITY:  Walnut Creek
;      STATE:  California
;      COUNTRY:  USA
```

```
;       ZIP:  94596
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS 5.0
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/607,538C
;       FILING DATE:  01-NOV-1990
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Viviana Amzel
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  CHPCC-004
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (510) 943-1931
;       TELEFAX:  (510) 943-1189
;       TELEX:  N.A.
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  218 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     FRAGMENT TYPE:
US-07-607-538C-2


  Query Match           100.0%;  Score 45;  DB 1;  Length 218;
  Best Local Similarity  100.0%;  Pred. No. 0.055;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
            |||||||||
Db     187 NLFETPILA 195


RESULT   5
US-08-162-402B-2
; Sequence 2, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  218 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-2


  Query Match           100.0%;  Score 45;  DB 2;  Length 218;
  Best Local Similarity  100.0%;  Pred. No. 0.055;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
            |||||||||
Db     187 NLFETPILA 195


RESULT   6
US-08-162-402B-6
; Sequence 6, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  387 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-6


  Query Match           100.0%;  Score 45;  DB 2;  Length 387;
  Best Local Similarity  100.0%;  Pred. No. 0.11;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 NLFETPILA 9
           |||||||||
Db     456 NLFETPILA 464


RESULT   7
US-08-162-402B-8
; Sequence 8, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213 622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  465 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-8


  Query Match             100.0%;  Score 45;  DB 2;  Length 465;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
           |||||||||
Db     434 NLFETPILA 442


RESULT   8
US-08-162-402B-10
; Sequence 10, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  160 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-10


  Query Match             82.2%;  Score 37;  DB 2;  Length 160;
  Best Local Similarity   77.8%;  Pred. No. 1.8;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
           |||||| :|
Db     126 NLFETPVEA 134


RESULT   9
US-08-162-402B-26
; Sequence 26, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
```

```
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Arnold, Virginia
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  FFF 38215
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  213-622-7700
;       TELEFAX:  213-489-4210
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  26:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  14 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
US-08-162-402B-26


  Query Match             71.1%;  Score 32;  DB 2;  Length 14;
  Best Local Similarity   71.4%;  Pred. No. 1.1;
  Matches    5;  Conservative    2,  Mismatches    0,  Indels    0  Gaps    0;

Qy       1 NLFETPI 7
           |||:||:
Db       2 NLFDTPL 8


RESULT  10
US-08-756-317-12
; Sequence 12, Application US/08756317
; Patent No. 5849894
;  GENERAL INFORMATION:
;     APPLICANT:  Clemente, Thomas E.
;     APPLICANT:  Kishore, Ganesh M.
;     APPLICANT:  Mitsky, Timothy A.
;     APPLICANT:  Stark, David M.
;     TITLE OF INVENTION:  Improved Rhodospirillum Rubrum
;     TITLE OF INVENTION:  Poly B Hydroxyalkanoate Synthase
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  TX
;       COUNTRY:  USA
;       ZIP:  77210-4433
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/756,317
;       FILING DATE:  25-NOV-1996
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/007,693
;       FILING DATE:  29-NOV-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Patterson, Melinda L.
;       REGISTRATION NUMBER:  33,062
;       REFERENCE/DOCKET NUMBER:  MOBT:008
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (713) 787-1400
;       TELEFAX:  (713) 787-1440
;   INFORMATION FOR SEQ ID NO:  12:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  590 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
US-08-756-317-12


  Query Match             71.1%;  Score 32;  DB 2;  Length 590;
  Best Local Similarity   62.5%;  Pred. No. 89;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPIL 8
           |:::||||
Db     239 NVYQTPIL 246


RESULT  11
US-08-822-445-12
; Sequence 12, Application US/08822445
; Patent No. 5952223
;  GENERAL INFORMATION:
;     APPLICANT:  Kaplan, Jerry
;     APPLICANT:  Perou, Charles
;     APPLICANT:  Moore, Karen
;     TITLE OF INVENTION:  COMPOSITIONS FOR THE DIAGNOSIS
;     TITLE OF INVENTION:  AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME
;     NUMBER OF SEQUENCES:  22
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/822,445
;       FILING DATE:  21-MAR-1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Corcuzi, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  298/062 999
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-790-9090
;       TELEFAX:  212-869-8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  12:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3672 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  protein
US-08-822-445-12


  Query Match             71.1%;  Score 32;  DB 2;  Length 3672;
  Best Local Similarity   75.0%;  Pred. No. 7.7e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPIL 8
           |||||:||
Db     443 NLFETAVL 450


RESULT  12
US-09-396-540-12
; Sequence 12, Application US/09396540
; Patent No. 6310182
;  GENERAL INFORMATION:
;     APPLICANT:  Kaplan, Jerry
```

```
;    APPLICANT:  Perou, Charles
;    APPLICANT:  Moore, Karen
;    TITLE OF INVENTION:  COMPOSITIONS FOR THE DIAGNOSIS
;    TITLE OF INVENTION:  AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/396,540
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/822,445
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-062-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3672 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-09-396-540-12


  Query Match             71.1%;  Score 32;  DB 4;  Length 3672;
  Best Local Similarity   75.0%;  Pred. No. 7.7e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPIL 8
           ||||| :|
Db     443 NLFETAVL 450


RESULT  13
US-08-822-445-10
; Sequence 10, Application US/08822445
; Patent No. 5952223
;  GENERAL INFORMATION:
;    APPLICANT:  Kaplan, Jerry
;    APPLICANT:  Perou, Charles
;    APPLICANT:  Moore, Karen
;    TITLE OF INVENTION:  COMPOSITIONS FOR THE DIAGNOSIS
;    TITLE OF INVENTION:  AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/822,445
;      FILING DATE:  21-MAR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-062-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3801 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-822-445-10


  Query Match             71.1%;  Score 32;  DB 2;  Length 3801;
  Best Local Similarity   75.0%;  Pred. No. 8e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPIL 8
           ||||| :|
Db     443 NLFETAVL 450


RESULT  14
US-09-396-540-10
; Sequence 10, Application US/09396540
; Patent No. 6310182
;  GENERAL INFORMATION:
;    APPLICANT:  Kaplan, Jerry
;    APPLICANT:  Perou, Charles
;    APPLICANT:  Moore, Karen
;    TITLE OF INVENTION:  COMPOSITIONS FOR THE DIAGNOSIS
;    TITLE OF INVENTION:  AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/396,540
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/822,445
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-062-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3801 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
```

```
;    MOLECULE TYPE:  protein
US-09-396-540-10


  Query Match             71.1%;  Score 32;  DB 4;  Length 3801;
  Best Local Similarity   75.0%;  Pred. No. 8e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPIL 8
            ||||| :|
Db     443 NLFETAVL 450


RESULT  15
US-08-480-229C-2
; Sequence 2, Application US/08480229C
; Patent No. 5874562
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS-1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
:      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,229C
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8907-0026-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
,      TELEFAX:  (212) 869-8864/9741
;      TELEX.  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  85 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-480-229C-2


  Query Match             68.9%;  Score 31;  DB 2;  Length 85;
  Best Local Similarity   55.6%;  Pred. No. 15;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
            |:|| | :|
Db      67 NIFEKPFMA 75


Search completed:  September  5, 2002, 15:27:34
Job time:  319 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:26:29 ; Search time 178.43 Seconds
(without alignments)
6,310 Million cell updates/sec

Title: US-09-744-804-37
Perfect score: 45
Sequence: 1 NLFETPIHA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	100.0	9	21	AAY82842	Lactadherin (BA-46
2	45	100.0	217	16	AAR77253	HMFG 46 kDa antige
3	45	100.0	218	16	AAR77254	HMFG 46 kDa antige
4	45	100.0	387	16	AAR77252	HMFG 46 kDa antige
5	45	100.0	387	21	AAY94453	Human lactadherin
6	37	82.2	9	21	AAY82843	Lactadherin (BA 46
7	35	77.8	1284	22	ABB62713	Drosophila melanog
8	35	77.8	1835	14	AAR42839	Urea amidolyase.
9	34	75.6	171	21	AAR41545	Human ORFX ORF1309
10	34	75.6	196	21	AAB23032	Human putative mit
11	34	75.6	306	22	AAU44139	Propionibacterium
12	32	71.1	26	22	AAB62694	ABC1 protein exter
13	32	71.1	302	22	AAU46488	Propionibacterium
14	32	71.1	333	22	ABB69663	Drosophila melanog
15	32	71.1	590	22	AAE10893	Acinetobacter sp.
16	32	71.1	1013	21	AAB01247	Human HE6 receptor
17	32	71.1	1038	18	AAW36903	Human epididymis-s
18	32	71.1	1089	22	AAB62691	Human ABC1 protein
19	32	71.1	1144	22	AAD02184	Human ABC1 mutant
20	32	71.1	1144	22	AAD02185	Human ABC1 mutant
21	32	71.1	1525	22	AAD02187	Human ABC1 mutant
22	32	71.1	1531	18	AAW23597	Human LYST1 short
23	32	71.1	2001	18	AAW23596	Human LYST1 longer
24	32	71.1	2130	22	AAD02190	Human ABC1 mutant
25	32	71.1	2143	21	AAB38108	Human ABC1 cholest
26	32	71.1	2201	21	AAY79380	Human ATP binding
27	32	71.1	2201	22	AAE13021	Human ATP binding
28	32	71.1	2201	22	AAM50227	Human ATP binding
29	32	71.1	2259	21	AAB38107	Human ABC1 FBA-3 m
30	32	71.1	2260	21	AAB38106	Human ABC1 cholest
31	32	71.1	2261	21	AAB39082	Human ABC1 cholest
32	32	71.1	2261	21	AAB38104	Human ABC1 cholest
33	32	71.1	2261	21	AAB38105	Human ABC1 cholest
34	32	71.1	2261	21	AAB38109	Human ABC1 cholest
35	32	71.1	2261	21	AAB38110	Human ABC1 cholest
36	32	71.1	2261	21	AAB38111	Human ABC1 cholest
37	32	71.1	2261	21	AAB38112	Human ABC1 cholest
38	32	71.1	2261	21	AAB38113	Human ABC1 cholest
39	32	71.1	2261	21	AAB38114	Human ABC1 cholest
40	32	71.1	2261	21	AAB38115	Human ABC1 cholest
41	32	71.1	2261	21	AAB38116	Human ABC1 cholest
42	32	71.1	2261	21	AAB38117	Human ABC1 cholest
43	32	71.1	2261	22	AAL19022	Human A-1 binding
44	32	71.1	2261	22	AAM22228	Human ATP binding
45	32	71.1	2261	22	AAM24554	Human protein sig

ALIGNMENTS

RESULT 1
AAY82842
ID AAY82842 standard; peptide; 9 AA.
XX
AC AAY82842;
XX
DT 19-JUN-2000 (first entry)
XX
DE Lactadherin (BA-46) peptide fragment (tumour associated antigen).
XX
KW Tumour associated antigen peptide; TAA; cancer; carcinoma;
KW treatment; prevention; cure; anti tumour vaccine; metastases;
KW breast, bladder, prostate, pancreas, ovary, thyroid, colon;
KW stomach; carcinoma; MHC Class I; HLA-A2; human;
KW Major Histocompatibility Complex; uroplakin;
KW prostate specific antigen; prostate specific membrane antigen;
KW prostate acid phosphatase; mucin; lactadherin;
KW teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRIPTO-1.
XX
OS Homo sapiens.
XX
PN WO200006723-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-IL000417.
XX
PR 30-JUL-1998; 98IL-0125608.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (BIOT) BIO TECHNOLOGY GEN CORP.
XX
PI Eisenbach L, Carmon L, Tirosh B, Barkalm E, Pac A, Fridkin M;
PI Fitzer-attas C;

```
XX
DR   WPI; 2000-205463/18.
XX
PT   Tumor associated antigen peptides, especially derived from uroplakin,
PT   useful as vaccines to prevent or cure cancers including breast,
PT   bladder, prostate, pancreas, ovary, thyroid, colon and stomach
XX
PS   Claim 17; Page 99; 113pp; English.
XX
CC   Tumour associated antigen peptides (TAA) may be used for the
CC   treatment, prevention and cure of cancer or cancer metastases. The
CC   cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
CC   colon, stomach, head or neck cancer or a carcinoma. The tumour
CC   associated antigens are presentable to the immune system by HLA-A2
CC   molecules and are generally between 8 to 10 amino acids in length.
CC   The amino acids located at positions 2 and 9 of the tumour associated
CC   antigens are the anchor residues which participate in the binding to
CC   MHC class I molecules, more specifically HLA-A2. More tumour
CC   associated antigens are described in GENESEQ records AAY82806-Y82882.
CC   Those tumour associated antigens described in records AAY82806-Y82824
CC   and AAY82855 YY82869 are derived from Uroplakin, such as Uroplakin II,
CC   Uroplakin Ia, Uroplakin III and Uroplakin Ib. Those described in
CC   records AAY82825-Y82829 are derived from prostate specific antigen
CC   (PSA). Those described in records AAY82830-Y82835 are derived from
CC   prostate specific membrane antigen (PSMA). Those described in
CC   records YY82836-AAY82839 are derived from prostate acid phosphatase
CC   (PAP). Those described in records AAY82840-Y82846 are derived
CC   from Lactadherin (BA 46). Those described in records AAY82847-Y82854
CC   are derived from Mucin and those described in records AAY82871-Y82882
CC   are derived from Teratocarcinoma derived growth factor (CRIPTO-1).
XX
SQ   Sequence   9 AA;


  Query Match             100.0%;  Score 45;  DB 21;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
           |||||||||
Db       1 nlfetpila 9


RESULT   2
AAR77253
ID   AAR77253 standard; Protein; 217 AA.
XX
AC   AAR77253;
XX
DT   21-NOV-1995  (first entry)
XX
DE   HMFG 46 kDa antigen C-terminal region.
XX
KW   HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
KW   epithelium; tumor; breast cancer; monoclonal antibody; MAb.
XX
OS   Homo sapiens.
XX
PN   WO9515171-A.
XX
PD   08-JUN-1995.
XX
PF   05-DEC-1994;   94WO-US13967.
XX
PR   03-DEC-1993;   93US-0162402.
XX
PA   (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI   Ceriani RL,  Larocca DJ,  Peterson JA;
XX
DR   WPI; 1995-215151/28.
XX
PT   46 kD apparent molecular weight human milk fat globule antigen  -
PT   used in assays to determine the presence of a cancerous tumour of
PT   epithelial origin, and in a vaccine against neoplastic tumours
XX
PS   Claim 7; Page 41; 68pp; English.
XX
CC   A partial cDNA clone BA46 1 (AAQ91199) for the 46 kDa HMFG antigen,
CC   a major component of the apical surface of the normal breast
CC   epithelial cell, was obtained by PCR of cDNA from a lactating
CC   breast cDNA library.  The C-terminal region of the encoded protein
CC   showed 43% identity to corresp. regions of human Factor-V and 48%
CC   to Factor-VIII.
XX
SQ   Sequence   217 AA;


  Query Match             100.0%;  Score 45;  DB 16;  Length 217;
  Best Local Similarity   100.0%;  Pred. No. 0.37;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
           |||||||||
Db     186 nlfetpila 194


RESULT   3
AAR77254
ID   AAR77254 standard; Protein; 218 AA.
XX
AC   AAR77254;
XX
DT   21-NOV-1995  (first entry)
XX
DE   HMFG 46 kDa antigen partial sequence.
XX
KW   HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
KW   epithelium; tumor; breast cancer; monoclonal antibody; MAb.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   69..71
FT                   /label= N-glycosylation_site
FT   Modified-site   156..158
FT                   /label= N-glycosylation_site
FT   Modified-site   160..162
FT                   /label= N-glycosylation_site
FT   Modified-site   181..183
FT                   /label= N-glycosylation_site
XX
PN   WO9515171-A.
XX
PD   08-JUN-1995.
XX
PF   05-DEC-1994;   94WO-US13967.
XX
PR   03-DEC-1993;   93US-0162402.
XX
PA   (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI   Ceriani RL,  Larocca DJ,  Peterson JA;
XX
DR   WPI; 1995-215151/28.
DR   N-PSDB; AAQ91199.
XX
PT   46 kD apparent molecular weight human milk fat globule antigen  -
PT   used in assays to determine the presence of a cancerous tumour of
PT   epithelial origin, and in a vaccine against neoplastic tumours
XX
PS   Example 7; Page 38-40; 68pp; English.
XX
CC   A partial cDNA clone BA46 1 (AAQ91199) for the 46 kDa HMFG antigen,
CC   a major component of the apical surface of the normal breast
CC   epithelial cell, was obtained by PCR of cDNA from a lactating
```

```
CC   breast cDNA library.
XX
SQ   Sequence   218 AA;


  Query Match             100.0%;  Score 45;  DB 16;  Length 218;
  Best Local Similarity   100.0%;  Pred. No. 0.37;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0,  Gaps    0,

Qy      1 NLFETPILA 9
          |||||||||
Db    187 nlfetpila 195


RESULT   4
AAR77252
ID   AAR77252 standard; Protein; 387 AA.
XX
AC   AAR77252;
XX
DT   21-NOV-1995  (first entry)
XX
DE   HMFG 46 kDa antigen.
XX
KW   HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
KW   epithelium; tumor; breast cancer; monoclonal antibody; MAb.
XX
OS   Homo sapiens.
XX
PN   WO9515171-A.
XX
PD   08-JUN-1995.
XX
PF   05-DEC-1994;   94WO-US13967.
XX
PR   03-DEC-1993;   93US-0162402.
XX
PA   (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI   Ceriani RL,  Larocca DJ,  Peterson JA.
XX
DR   WPI; 1995-215151/28.
DR   N-PSDB; AAQ91198.
XX
PT   46 KD apparent molecular weight human milk fat globule antigen
PT   used in assays to determine the presence of a cancerous tumour of
FT   epithelial origin, and in a vaccine against neoplastic tumours
XX
PS   Claim 6; Page 46-47; 68pp; English
XX
CC   A complete cDNA sequence for the 46 kDa HMFG antigen, a major
CC   component of the apical surface of the normal breast epithelial
CC   cell, was obtained by PCR and RACE methods.  cDNA clones can be
CC   used to prepare MAbs for use in immunotherapy, immunohistopathology,
CC   prognosis, diagnosis, imaging and therapy.  Recombinant antigen can
CC   be expressed in prokaryotic or (glycosylated) in eukaryotic cells.
XX
SQ   Sequence   387 AA;


  Query Match             100.0%;  Score 45;  DB 16;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.68;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NLFETPILA 9
          |||||||||
Db    356 nlfetpila 364


RESULT   5
AAY94453
ID   AAY94453 standard; Protein; 387 AA.
XX
AC   AAY94453;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Human lactadherin protein.
XX
KW   Human; lactadherin; MGF-E8; anti-tumour; immune response;
KW   exosome; dendritic cell.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Secretion_signal
FT   Protein         24..387
FT                   /label= Lactadherin
FT   Binding-site    46..48
FT                   /label= Integrin_binding_site
XX
PN   EP1004664-A1.
XX
PD   31-MAY-2000.
XX
PF   24-NOV-1998;   98EP-0402925.
XX
PR   24-NOV-1998;   98EP-0402925.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (CURI-) INST CURIE.
XX
DR   WPI; 2000-452597/31.
DR   N-PSDB; AAA27140.
XX
PT   Chimeric isolated (human) lactadherin polypeptide that functions as an
PT   adaptor of cross priming to eliminate pathogenic antigens, e.g. in
PT   tumors -
XX
PS   Example 3; Page 12; 20pp; English.
XX
CC   Lactadherin protein was found in exosomes produced by dendritic cells.
CC   The protein is involved in the phagocytosis of particulate antigens by
CC   dendritic cells. Exosomes produced by dendritic cells exposed to
CC   tumour antigens induce potent immune responses. Lactadherin or variants
CC   of it may be used in the mediation of an immune response. Variants of
CC   lactadherin may be used for inhibition and/or stimulation of the
CC   cross priming of antigens and stimulation of the phagocytosis of
CC   antigens by dendritic cells. Compositions derived from lactadherin can
CC   also be used to monitor an immune response, more specifically a CTL
CC   (cytotoxic T-lymphocyte) response and also to produce CTLs specific
CC   for a selected antigen. The present sequence is the human
CC   lactadherin protein.
XX
SQ   Sequence   387 AA;


  Query Match             100.0%;  Score 45;  DB 21;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.68;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NLFETPILA 9
          |||||||||
Db    356 nlfetpila 364


RESULT   6
AAY82843
ID   AAY82843 standard; peptide; 9 AA.
XX
AC   AAY82843;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Lactadherin (BA-46) peptide fragment (tumour associated antigen).
```

XX
KW Tumour associated antigen peptide; TAA; cancer; carcinoma;
KW treatment; prevention; cure; anti-tumour vaccine; metastases;
KW breast; bladder; prostate; pancreas; ovary; thyroid; colon;
KW stomach; carcinoma; MHC Class I; HLA-A2; human;
KW Major Histocompatibility Complex; uroplakin;
KW prostate specific antigen; prostate specific membrane antigen;
KW prostate acid phosphatase; mucin; lactadherin;
KW teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRIPTO-1
XX
OS Homo sapiens.
XX
PN WO200006723-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-IL00417.
XX
PR 30-JUL-1998; 98IL-0125608.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (BIOT-) BIO-TECHNOLOGY GEN CORP.
XX
PI Eisenbach L, Carmon L, Tirosh B, Bar-haim E, Paz A, Fridkin M;
PI Fitzer-attas C;
XX
DR WPI; 2000-205463/18.
XX
PT Tumor associated antigen peptides, especially derived from uroplakin,
PT useful as vaccines to prevent or cure cancers including breast,
PT bladder, prostate, pancreas, ovary, thyroid, colon and stomach
XX
PS Claim 17; Page 100; 113pp; English.
XX
CC Tumour associated antigen peptides (TAA) may be used for the
CC treatment, prevention and cure of cancer or cancer metastases. The
CC cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
CC colon, stomach, head or neck cancer or a carcinoma. The tumour
CC associated antigens are presentable to the immune system by HLA-A2
CC molecules and are generally between 8 to 10 amino acids in length.
CC The amino acids located at positions 2 and 9 of the tumour associated
CC antigens are the anchor residues which participate in the binding to
CC MHC class I molecules, more specifically HLA-A2. More tumour
CC associated antigens are described in GENESEQ records AAY82806-Y82882.
CC Those tumour associated antigens described in records AAY82806-Y82824
CC and AAY82855-YY82869 are derived from Uroplakin, such as Uroplakin II,
CC Uroplakin Ia, Uroplakin III and Uroplakin Ib. Those described in
CC records AAY82825-Y82829 are derived from prostate specific antigen
CC (PSA). Those described in records AAY82830-Y82835 are derived from
CC prostate specific membrane antigen (PSMA). Those described in
CC records YY82836 AAY82839 are derived from prostate acid phosphotase
CC (PAP). Those described in records AAY82840-Y82846 are derived
CC from Lactadherin (BA-46). Those described in records AAY82847-Y82854
CC are derived from Mucin and those described in records AAY82871-YY82882
CC are derived from Teratocarcinoma derived growth factor (CRIPTO-1).
XX
SQ Sequence 9 AA;

Query Match 82.2%; Score 37; DB 21; Length 9;
Best Local Similarity 77.8%; Pred. No. 6.4e+05;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLFETPILA 9
|||||| :
Db 1 nlfetpvea 9

RESULT 7
ABB62713
ID ABB62713 standard; Protein; 1284 AA.
XX
AC ABB62713;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 14931.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-655869/75.
DR N-PSDB; ABL06816.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 14931; 21pp - Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840 ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1284 AA;

Query Match 77.8%; Score 35; DB 22; Length 1284;
Best Local Similarity 66.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLFETPILA 9
|:|| ||:|
Db 154 nvfevpila 162

RESULT 8
AAR42839
ID AAR42839 standard; Protein; 1835 AA.
XX
AC AAR42839;
XX
DT 10-MAY-1994 (first entry)
XX
DE Urea amidolyase.
XX
KW Urea amidolyase; URL; yeast; recombinant plasmid.
XX
OS Saccharomyces cerevisiae.
XX
PN JP05244959-A.
XX
PD 24-SEP-1993.

XX
PF 05-MAR-1992; 92JP-0084541.
XX
PR 05-MAR-1992; 92JP-0084541.
XX
PA (TOYM) TOYOBO KK.
XX
DR WPI; 1993-338925/43
DR N-PSDB; AAQ49460.
XX
PT DNA having the genetic information of urea amidolyase originated
PT from Saccharomyces yeast can be used to prepare high purity
PT urea amidolyase by culturing the transformant comprising the DNA
XX
PS Claim 2, Page 16-17, 17pp, Japanese.
XX
CC This sequence represents a protein which has urea amidolyase (URL)
CC activity and is derived from yeast. The DNA encoding this protein
CC may be used within a recombinant plasmid for the production of highly
CC pure URL.
XX
SQ Sequence 1835 AA;

Query Match 77.8%; Score 35; DB 14; Length 1835;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 3 PEIPILA 9
||||||| |
Db 1418 PETPILA 1424

RESULT 9
AAB41545
ID AAB41545 standard; Protein; 171 AA.
XX
AC AAB41545;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1309 polypeptide sequence SEQ ID NO:2618.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder, asthma,
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease, coagulation,
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR N-PSDB; AAC75754.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11, Page 1863-1864, 5507pp, English.
XX
CC AAC74445 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic, hepatotropic, vulnerary;
CC antipsoriatic, antiparkinsonian; nootropic, neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant, cardiant, thrombolytic, coagulant, vasotropic;
CC antidiabetic, hypotensive, dermatological, immunosuppressive;
CC antiinflammatory, antibacterial, antiviral, antifungal, antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 171 AA;

Query Match 75.6%; Score 34; DB 21; Length 171;
Best Local Similarity 66.7%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLFEIPILA 9
|: ||||:
Db 27 nimetpils 35

RESULT 10
AAB23032
ID AAB23032 standard; Protein; 196 AA.
XX
AC AAB23032;
XX
DT 16-JAN-2001 (first entry)
XX
DE Human putative mitochondrial protein, SECX 2982339.
XX
KW SECX protein, human, secreted, membrane-associated; cancer;
KW proliferation regulator; differentiation regulator; non-malignant tumour;
KW immune disorder; autoimmune disease, transplant rejection, allergy; AIDS;
KW infection; inflammatory disorder; arthritis; haematopoietic disorder;
KW skin disorder; cardiovascular disorder, atherosclerosis, restenosis,
KW neurological disease, Alzheimer's disease, trauma, wounding;
KW spinal cord injury; skeletal disorder; cytostatic; immunosuppressive;
KW anti-HIV; antiinflammatory; antiarthritic; antiarteriosclerotic;
KW neuroprotective; vulnerary; antiallergic; antimicrobial; cardiant;
KW dermatological; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200053742-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06280.

```
XX
PR   09-MAR-1999;   99US-0123667.
PR   08-MAR-2000;   2000US-0123667.
XX
PA   (CURA-) CORIXA CORP.
XX
PI   Shimkets RA;
XX
DR   WPI; 2000-594318/56
DR   N-PSDB; AAA93619.
XX
PT   Novel human membrane associated or secreted polypeptides and
PT   polynucleotides useful for diagnosis, prevention and treatment of
PT   pathological states such as cancer, immune, cardiovascular and
PT   neurological disorders -
XX
PS   Claim 1; Fig 4; 151pp; English.
XX
CC   Sequences AAB23029 B23048 represent human SECX proteins. The SECX
CC   proteins of the invention are either secreted or membrane associated
CC   proteins and act as regulator of cellular proliferation and
CC   differentiation. SECX proteins or nucleotides are useful for diagnosing
CC   the presence of, or predisposition to, a disease associated with altered
CC   levels of SECX proteins and nucleotides. The SECX proteins are also
CC   useful to screen compounds that modulate SECX activity or expression. The
CC   interaction of a SECX protein with other cellular proteins may be useful
CC   to modulate the activity of a partner protein, cellular proliferation,
CC   cellular differentiation and cell survival. SECX nucleotides are useful
CC   for the recombinant expression of SECX protein, and may be used detect
CC   SECX mRNA or genetic lesions in the SECX gene. They may also be used to
CC   modulate SECX expression (e.g., using antisense oligonucleotides). SECX
CC   nucleic acid sequences are also useful for identifying a cell or tissue
CC   type in a biological sample, and in forensic biology. SECX primers or
CC   probes are useful for detecting the presence of SECX nucleotides and for
CC   screening tissue cultures for contamination. Diseases that may be treated
CC   or prevented using SECX proteins or nucleotides include cancer (e.g.,
CC   colorectal carcinoma, prostate cancer), benign tumours, immune disorders
CC   (including autoimmune diseases, transplant rejection, allergies, AIDS),
CC   infections, inflammatory disorders, arthritis, haematopoietic disorders,
CC   skin disorders, cardiovascular disorders, atherosclerosis, restenosis,
CC   neurological diseases (e.g., Alzheimer's disease), trauma (e.g.
CC   surgical or traumatic wounds, spinal cord injury), and skeletal
CC   disorders.
XX
SQ   Sequence   196 AA;


  Query Match             75.6%;  Score 34;  DB 21;  Length 196;
  Best Local Similarity   75.0%;  Pred. No. 51;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPIL 8
             ||| ||:|
Db      51 nltstpll 58


RESULT  11
AAU44139
ID   AAU44139 standard; Protein; 306 AA.
XX
AC   AAU44139;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #5035.
XX
KW   SAPHO syndrome; synovitis, acne, pustulosis, hypertosis, osteomyelitis;
KW   uveitis; endophthalmitis, bone, joint, central nervous system, ELISA,
KW   inflammatory lesion, acne vulgaris, enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PP   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'Italien J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59521.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Claim 3; SEQ ID NO 5334; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   306 AA;


  Query Match             75.6%;  Score 44;  DB 22;  Length 306;
  Best Local Similarity   66.7%;  Pred. No. 82;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
             |: ||||:
Db      40 nimetpils 48


RESULT  12
AAB62694
ID   AAB62694 standard; peptide; 26 AA.
XX
AC   AAB62694;
XX
DT   06-AUG-2001  (first entry)
XX
DE   ABC1 protein external domain TM5 TM6 fragment (residues 795 820).
XX
KW   ABC1, antilipemic, cholesterol, inhibitor, low density lipoprotein; LDL.
XX
OS   Homo sapiens.
XX
PN   WO200132184-A2.
XX
```

PD 10-MAY-2001.
XX
PF 01-NOV-2000; 2000WO-US30109.
XX
PR 01-NOV-1999; 99US-0162803.
PR 30-JUN-2000; 2000US-0215564.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Attie AD, Cook M, Gray Keller MP, Hayden MR, Pimstone S;
PI Brooks-Wilson A;
XX
DR WPI; 2001-335779/35.
XX
PT New method for inhibiting cholesterol uptake in the gut comprises
PT administration of an inhibitor of an ABC1 protein -
XX
PS Disclosure; Page 9; 41pp; English.
XX
CC The invention relates to a new method for inhibiting cholesterol uptake
CC in the gut that comprises administration of an inhibitor of an ABC1
CC protein. The method is useful for: lowering levels of LDL (low density
CC lipoprotein) cholesterol by reducing the activity of ABC1 protein in the
CC intestinal cells and the abundance of the ABC1 protein in the individual
CC by inhibiting the activity of the protein; identifying drugs that can
CC lower serum cholesterol and LDL levels comprises assaying the drug to
CC test if it can bind to an ABC1 protein; testing LDL cholesterol lowering
CC agents; and for modulation of ABC1 biological activity. Sequences
CC AAB62692-97 represent predicted external domain of ABC1 protein.
XX
SQ Sequence 26 AA;

Query Match 71.1%; Score 32; DB 22; Length 26;
Best Local Similarity 71.4%; Pred. No. 15;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPI 7
||||:|:
Db 14 nlfespv 20

RESULT 13
AAU46488
ID AAU46488 standard; Protein; 302 AA
XX
AC AAU46488;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #7384.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59542.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 7683; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC down-regulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 302 AA;

Query Match 71.1%; Score 32; DB 22; Length 302;
Best Local Similarity 77.8%; Pred. No. 2e-02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NLFETPILA 9
||||| ||
Db 176 nlfetqtla 184

RESULT 14
ABB69663
ID ABB69663 standard; Protein; 333 AA.
XX
AC ABB69663;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 35781.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656890/75.
DR N-PSDB; ABL13766.

```
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 35781; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The Sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   333 AA;


  Query Match             71.1%;  Score 32;  DB 22;  Length 333;
  Best Local Similarity   85.7%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LFETPIL 8
           |||||:|
Db     228 lfetpml 234


RESULT  15
AAE10893
ID   AAE10893 standard; Protein; 590 AA.
XX
AC   AAE10893;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Acinetobacter sp. RA3849 polyhydroxybutyrate (PHB) synthase.
XX
KW   Polyhydroxybutyrate synthase; PHB; sphingan; food product; dessert gel;
KW   jelly; jam; beverage; dairy product; gelling agent; rheological modifier;
KW   industrial application.
XX
OS   Acinetobacter sp. RA3849.
XX
PN   WO200164897-A2.
XX
PD   07-SEP-2001.
XX
PF   02-MAR 2001; 2001WO-US07010.
XX
PR   02-MAR-2000; 2000US-186433P.
XX
PA   (KELC ) CP KELCO US INC.
XX
PI   Bower S,  Burke E,  Harding N,  Patel YN,  Schneider JC,  Meissner D;
PI   Morrison N,  Bezanson R;
XX
DR   WPI; 2001-589870/66.
XX
PT   Mutant strain of Sphingomonas species useful for producing sphingan,
PT   has a mutation in the gene encoding a protein involved in
PT   polyhydroxybutyrate synthesis -
XX
PS   Example 1; Page 84-86; 98pp; English.
XX
CC   The invention relates to mutant strains of Sphingomonas species which
CC   have a mutation in the gene encoding a protein involved in internal
CC   storage polymer polyhydroxybutyrate (PHB) synthesis that allows the
CC   mutant strains to produce PHB-deficient sphingans. Sphingans are capsular
CC   polysaccharides secreted by bacteria of the genus Sphingomonas. The
CC   invention also relates to a process for preparing clarified sphingans
CC   which are useful as gelling agents in a variety of food applications
CC   for improving the taste, texture, stability and appearance of food
CC   products such as dessert gels, confectionery jellies, jams, dairy
CC   products, beverages, films and coatings. The sphingans are also useful
CC   as rheological modifier in industrial applications such as oil field
CC   drilling and cementitious systems. The present sequence is
CC   Acinetobacter sp. strain RA3849 poly(3-hydroxybutyrate (PHB) synthase which
CC   is used for the generation of Sphingomonas elodea phaC fragment. phaC DNA
CC   fragment encodes polyhydroxybutyrate (PHB) synthase protein.
XX
SQ   Sequence   590 AA;


  Query Match             71.1%;  Score 32;  DB 22;  Length 590;
  Best Local Similarity   62.5%;  Pred. No. 4.1e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPIL 8
             |:::||||
Db     239 nvyqtpil 246


Search completed: September  5, 2002, 15:26:40
Job time: 370 sec
```

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:28:50 ; Search time 69.78 Seconds
(without alignments)
12.393 Million cell updates/sec

Title: US-09-744-804-37
Perfect score: 45
Sequence: 1 NLFETPILA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	45	100.0	218	2	A47285	milk fat globule p
2	36	80.0	612	2	T11830	NADH dehydrogenase
3	36	80.0	669	2	T48466	hypothetical prote
4	36	80.0	705	2	T48464	hypothetical prote
5	35	77.8	409	2	T11743	pP47 protein - pig
6	35	77.8	1283	2	T13799	neurexin IV - frui
7	35	77.8	1835	2	S46082	urea carboxylase (
8	34	75.6	142	2	A12479	hypothetical prote
9	34	75.6	250	2	A70554	probable dehydroge
10	34	75.6	310	2	A12848	ribose phosphate p
11	34	75.6	318	2	Q97625	prsa (AE000059) [i
12	33	73.3	304	2	A11303	dihydroorotase deh
13	33	73.3	304	2	A11675	dihydroorotase deh
14	33	73.3	363	2	I37716	actin-interacting
15	33	73.3	401	2	S65138	glycoprotein antig
16	33	73.3	427	2	S74211	PAS-6/7 protein pr
17	33	73.3	464	2	T16889	hypothetical prote
18	33	73.3	535	1	S76953	protein kinase (EC
19	33	73.3	572	2	T32636	hypothetical prote
20	33	73.3	658	2	T40107	hypothetical 57.9
21	33	73.3	664	2	T23325	hypothetical prote
22	32	71.1	121	2	G71111	hypothetical prote
23	32	71.1	173	2	G82062	hypothetical prote
24	32	71.1	208	2	B71031	hypothetical prote
25	32	71.1	212	2	A11836	hypothetical prote
26	32	71.1	255	2	H83326	probable short cha
27	32	71.1	268	2	G97394	undecaprenol kinas
28	32	71.1	268	2	AF2612	bacitracin resista
29	32	71.1	268	2	AH3541	bacitracin resista
30	32	71.1	278	2	F87289	molybdenum transpo
31	32	71.1	394	2	D82725	nicotinate phospho
32	32	71.1	398	2	F84099	nicotinate phospho
33	32	71.1	399	2	E83040	nicotinate phospho
34	32	71.1	417	2	T50165	probable polyipoly
35	32	71.1	425	2	F81706	conserved hypothet
36	32	71.1	439	2	H84699	sodium dependent c
37	32	71.1	494	2	AC0133	probable permease
38	32	71.1	527	2	S62484	hypothetical prote
39	32	71.1	530	1	T41892	probable dihydroli
40	32	71.1	532	1	WZBEF5	59K transcription
41	32	71.1	553	1	H70786	probable dihydroli
42	32	71.1	590	2	I39685	poly(n-3)pufa elki
43	32	71.1	670	2	C84496	probable replicati
44	32	71.1	757	2	C89473	protein F52D2.7 [i
45	32	71.1	779	2	S64993	hypothetical prote

ALIGNMENTS

RESULT 1
A47285
milk fat globule protein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-May-1999
C;Accession: A47285
R;Larocca, D., Peterson, J.A., Urrea, R., Kuniyoshi, J., Bistrain, A.M., Ceriani, R.L.
Cancer Res. 51, 4994-4998, 1991
A;Title: A M-r 46,000 human milk fat globule protein that is highly expressed in huma
A;Reference number: A47285; MUID:91371351
A;Accession: A47285
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-218 <LAR>
A;Cross-references: GB:S56151; NID:g295346; PIDN:AAB19771.1; PID:g295347
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
F;1-54/Domain: discoidin I amino-terminal homology (fragment) <DN1>
F;60-218/Domain: discoidin I amino-terminal homology <DN2>

Query Match 100.0%; Score 45; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 0.067;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPILA 9
||||||||| |
Db 187 NLFETPILA 195

RESULT 2
T11830
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Atlantic cod mitochondrion
C;Species: mitochondrion Gadus morhua (Atlantic cod)
C;Date: 16-Jul-1994 #sequence_revision 16-Jul-1994 #text_change 30-Jun-2000
C;Accession: T11830
R;Johansen, S., Bakke, I.
Mol. Marine Biol. Biotechnol. 5, 203-214, 1996
A;Title: The complete mitochondrial DNA sequence of Atlantic cod, Gadus morhua. Relev
A;Reference number: Z17351; MUID:96414925
A;Accession: T11830
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-612 <JOH>
A;Cross-references: EMBL:X99772; PIDN:CAA68115.1
A;Experimental source: Norwegian coastal stock (NC-1)
C;Genetics:
A;Genome: mitochondrion
A;Genetic code: SGC1
A;Note: ND5
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C;Keywords: #membrane-associated complex; #1 chondrion; NAD; oxidati + phosphorylatio

Query Match 80.0%; Score 36; DB 2; Length 612;
Best Local Similarity 75.0%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPIL 8
||:|||:|
Db 479 NLFKTPVL 486

RESULT 4
T48466
hypothetical protein T1E3.60 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 28-Jul-2000
C;Accession: T48466
R;Bevan, M.; Terryn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, F.; De Clerck, R.; De
ewes, H.W.; Rudd, S.; Tzafrir, I.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24491
A;Accession: T48466
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-669 <BEV>
A;Cross-references: EMBL:AL162972
A;Experimental source: cultivar Columbia; BAC clone T1E3
C;Genetics:
A;Map position: 5
A;Introns: 72/1; 226/1; 283/1; 307/1; 322/1; 465/3
A;Note: T1E3.60
C;Superfamily: Arabidopsis thaliana hypothetical protein T1E3.60

Query Match 80.0%; Score 36; DB 2; Length 669;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPIL 8
||:|||:|
Db 112 NLYETPLL 119

RESULT 4
T48464
hypothetical protein T1E3.40 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 28-Jul-2000
C;Accession: T48464
R;Bevan, M.; Terryn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, F.; De Clerck, R.; De
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24491
A;Accession: T48464
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-705 <BEV>
A;Cross-references: EMBL:AL162972
A;Experimental source: cultivar Columbia; BAC clone T1E3
C;Genetics:
A;Map position: 5
A;Introns: 59/1; 286/1; 457/1; 466/3
A;Note: T1E3.40

Query Match 80.0%; Score 36; DB 2; Length 705;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPIL 8
||:|||:|
Db 99 NLYETPLL 106

RESULT 5
T11743
pP47 protein - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T11743
R;Ensslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toe
Biol Reprod 58, 1057-1064, 1998
A;Title: Molecular cloning and characterization of P47, a novel boar sperm-associated
A;Reference number: Z17325; MUID:98206817
A;Accession: T11743
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-409 <ENS>
A;Cross-references: EMBL:Y11683; NID:g2652927; PIDN:CAA72479.1; PID:g2652928
A;Experimental source: testis
C;Function:
A;Description: may be involved in membrane remodelling and/or function as a zona pellu
C;Superfamily: unassigned EGF related proteins; EGF homology
F;6-40/Domain: EGF homology <EGF>

Query Match 77.8%; Score 35; DB 2; Length 409;
Best Local Similarity 75.0%; Pred. No. 18;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLFETPIL 8
|:|||:|
Db 378 NMFETPFL 385

RESULT 6
T13799
neurexin IV - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13799
R;Baumgartner, S.; Littleton, J.T.; Broadie, K.; Bhat, M.A.; Harbecke, R.; Lengyel, J
Cell 87, 1059-1068, 1996
A;Title: A Drosophila neurexin is required for septate junction and blood-nerve barri
A;Reference number: Z17762; MUID:97134213
A;Accession: T13799
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1283 <BAU>
A;Cross-references: EMBL:X86685; NID:g1518220; PIDN:CAA60483.1; PID:g1518221
C;Genetics:
A;Gene: nrx
A;Cross-references: FlyBase:FBgn0013997
C;Function:
A;Description: is required for NRX in the formation of septate junction septa and int

Query Match 77.8%; Score 35; DB 2; Length 1283;
Best Local Similarity 66.7%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLFETPILA 9
|:|| ||:|
Db 153 NVFEVPIIA 161

RESULT 7
S46082
urea carboxylase (EC 6.3.4.6) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YBR1449; protein YBR208c; urea amidolyase
C;Species: Saccharomyces cerevisiae
C;Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 11-Jan-2002
C;Accession: S46082; S46081; S46080; S31341; S34930; S34943
R;Rieger, M.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45734
A;Accession: S46082

A;Molecule type: DNA
A;Residues: 1-893 <PIE>
A;Cross-references: EMBL:Z36077, MIPS:YBR208c
A;Experimental source: strain S288C
R;Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45927
A;Accession: S46081
A;Molecule type: DNA
A;Residues: 873-1835 <FEL>
A;Cross-references: EMBL:Z36077; MIPS:YBR208c
A;Experimental source: strain S288C
R;Bussereau, F.; Demolis, N.; Jacquet, M.; Mallet, L.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S46054
A;Accession: S46080
A;Molecule type: DNA
A;Residues: 1487-1835 <BUS>
A;Cross-references: EMBL:Z36077; MIPS:YBR208c
A;Experimental source: strain S288C
R;Genbauffe, F.S.; Cooper, T.G.
DNA Seq. 2, 19-42, 1991
A;Title: The urea amidolyase (DUR1,2) gene of Saccharomyces cerevisiae.
A;Reference number: S31341; MUID:92199240
A;Accession: S31341
A;Molecule type: DNA
A;Residues: 1-95,'P',97-255,257-258,'N',259-458,'M',460-829,'K',831-1394,'H',1396-1835
A;Cross-references: EMBL:M64926; NID:g173121; PIDN:AAC41643.1; PID:g173122
R;Bussereau, F.; Mallet, L.; Caillon, L.; Jacquet, M.
Yeast 9, 797-806, 1993
A;Title: Yeast Sequencing Reports. A 12.8 kb segment, on the right arm of chromosome II
A;Reference number: S34925; MUID:93377417
A;Accession: S34930
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1487-1835 <BU2>
A;Cross-references: EMBL:Z21487; NID:g311665; PIDN:CAA79685.1; PID:g311684
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:DUR1,2
A;Cross-references: SGD:S0000412; MIPS:YBR208c
A;Map position: 2P
C;Superfamily: lipoyl/biotin-binding homology; biotin carboxylase homology
C;Keywords: ATP; biotin binding; ligase; P loop; purine nucleotide binding; transmembran
F;122-129/Region: nucleotide-binding motif A (P-loop)
F;163-179/Domain: transmembrane #status predicted <TM1>
F;209-225/Domain: transmembrane #status predicted <TM2>
F;411-427/Domain: transmembrane #status predicted <TM3>
F;635-1082/Domain: biotin carboxylase homology; BCH
F;1141-1157/Domain: transmembrane #status predicted <TM4>
F;1582-1598/Domain: transmembrane #status predicted <TM5>
F;1759-1832/Domain: lipoyl/biotin-binding; homology; LBD
F;128/Binding site: ATP/GTP (Lys) #status predicted
F;1798/Binding site: biotin (Lys) (covalent) #status predicted

  Query Match 77.8%; Score 35; DB 2; Length 1835;
  Best Local Similarity 100.0%; Pred. No. 1e+02;
  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FETPILA 9
|||||||
Db 1418 FETPILA 1424

RESULT 8
AI2479
hypothetical protein alr7017 [imported] - Anabaena sp. (strain PCC 7120) plasmid pCC7120
C;Species: Anabaena sp.
A;Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C;Date: 14-Dec 2001 #sequence_revision 14-Dec 2001 #text_change 11 Jan 2002
C;Accession: AI2479
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N., Shimpo, S., Sugimoto, M., Takazawa, M., Yamada, M., Yasuda, M., Tabata
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2479
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-142 <KUR>
A;Cross-references: GB:BA000020; PIDN:BAB78101.1; PID:g17135555; GSPDB:GN00180
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr7017
A;Genome: plasmid

  Query Match 75.6%; Score 34; DB 2; Length 142;
  Best Local Similarity 75.0%; Pred. No. 8.8;
  Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLFETPIL 8
|||| |:|
Db 65 NLFRIPVL 72

RESULT 9
A70554
probable dehydrogenase - Mycobacterium tuberculosis (strain H37Rv)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jun-2000
C;Accession: A70554
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
, Eiglmeier, K.; Gas, S.; Barry, C.E. 3rd; Tekaia, F.; Badcock, K.; Basham, D.; Brown,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: A70554
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-250 <COL>
A;Cross-references: GB:Z95584; GB:AL123456; NID:g2261774; PIDN:CAB09032.1; PID:g21171
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1144
C;Superfamily: ribitol dehydrogenase; short chain alcohol dehydrogenase homology
F;6-187/Domain: short chain alcohol dehydrogenase homology; SADH

  Query Match 75.6%; Score 34; DB 2; Length 250;
  Best Local Similarity 75.0%; Pred. No. 17;
  Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LFETPILA 9
||:|||:|
Db 188 LFDTPLLA 195

RESULT 10
AI2848
ribose-phosphate pyrophosphokinase [imported] - Agrobacterium tumefaciens (strain C58
C;Species: Agrobacterium tumefaciens
C;Date: 11 Jan 2002 #sequence_revision 11-Jan-2002 #text_change 01 Feb 2002
C;Accession: AI2848
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
eraje, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl
, Karp, P., Romero, P., Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AI2848

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-310 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL43207.1; PID:g17748688; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: prsA
A;Map position: circular chromosome
C;Superfamily: ribose-phosphate pyrophosphokinase catalytic chain

Query Match 75.6%; Score 34; DB 2; Length 310;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NLFETPILA 9
|||| ||||
Db 141 NLFAAPILA 149

RESULT 11
G97625
prsA (AF006059) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jun-2002
C;Accession: G97625
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollum, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, R.
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
A;Reference number: A97359; PMID:11743194
A;Accession: G97625
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-318 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87960.1; PID:g15157366; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_4031
A;Map position: circular chromosome
C;Superfamily: ribose phosphate pyrophosphokinase catalytic chain

Query Match 75.6%; Score 34; DB 2; Length 318;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NLFETPILA 9
|||| ||||
Db 149 NLFAAPILA 157

RESULT 12
AI1303
dihydroorotase dehydrogenase homolog pyrD [imported] - Listeria monocytogenes (strain EG
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AI1303
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mavok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AI1303
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-304 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99911.1; PID:g16411287; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: pyrD
C;Superfamily: dihydroorotate oxidase

Query Match 73.3%; Score 33; DB 2; Length 304;
Best Local Similarity 85.7%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FETPILA 9
||||:|
Db 92 FETPIIA 98

RESULT 13
AI1675
dihydroorotase dehydrogenase homolog pyrD [imported] - Listeria innocua (strain Clip1
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AI1675
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mavok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AI1675
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-304 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC97172.1; PID:g16411448; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: pyrD
C;Superfamily: dihydroorotate oxidase

Query Match 73.3%; Score 33; DB 2; Length 304;
Best Local Similarity 85.7%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FETPILA 9
||||:|
Db 92 FETPIIA 98

RESULT 14
T37716
actin-interacting protein bud6/Aip3p protein SPAC15E1.01 [imported] - fission yeast (
C;Species: Schizosaccharomyces pombe
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C;Accession: T37716
R;Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Lyne, M.H.
submitted to the EMBL Data Library, August 1999
A;Reference number: Z21739
A;Accession: T37716
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-363 <MUR>
A;Cross-references: EMBL:AL109770; PIDN:CAB52420.1; GSPDB:GN00066; SPDB:SPAC15E1.01
A;Experimental source: strain 972h-; cosmid c15E1
C;Genetics:
A;Gene: SPAC15A10.15, SPDB.SPAC15E1.01
A;Map position: 1

Query Match 73.3%; Score 33; DB 2; Length 364;
Best Local Similarity 62.5%; Pred. No. 42;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPIL 8
|::||||:
Db 282 NIYETPIV 289

RESULT 15
S65138
glycoprotein antigen MGP57/53, mammary gland - bovine (fragment)
N;Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E8 h
C;Species: Bos primigenius taurus (cattle)
C;Date: 28-Oct-1996 #sequence_revision 13 Mar 1997 #text_change 07-Aug-1998
C;Accession: S65138; G48394
R;Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A;Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal an
A;Reference number: S65138; MUID:96125736
A;Accession: S65138
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-401 <AOK>
R;Mather, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A;Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea pig
II-like sequences.
A;Reference number: A48394; MUID:93250576
A;Accession: G48394
A;Status: preliminary
A;Molecule type: protein
A;Residues: 207-220 <MAT>
A;Experimental source: milk
A;Note: sequence extracted from NCBI backbone (NCBIP:131457)
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
C;Keywords: glycoprotein
F;1-32/Domain: EGF homology (fragment) <EG1>
F;40-79/Domain: EGF homology <EG2>
F;82-239/Domain: discoidin I amino-terminal homology <DN1>
F;243-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match 73.3%; Score 33; DB 2; Length 401;
Best Local Similarity 66.7%; Pred. No. 48;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 NLFETPILA 9
           |:|||  |
Db     370 NIFETPFQA 378
```

Search completed: September 5, 2002, 15:28:51
Job time: 356 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:40 ; Search time 43.93 Seconds
(without alignments)
10.252 Million cell updates/sec

Title: US-09-744-804-37
Perfect score: 45
Sequence: 1 NLFETPILA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	100.0	387	1	MFGM_HUMAN	Q08431 homo sapien
2	36	80.0	612	1	NUSM_GADMO	P55782 gadus morhu
3	35	77.8	409	1	MFGM_PIG	P79385 sus scrofa
4	35	77.8	1284	1	NRX4_DROME	Q94887 drosophila
5	35	77.8	1835	1	DUR1_YEAST	P32528 saccharomyc
6	33	73.3	427	1	MFGM_BOVIN	Q95114 bos taurus
7	33	73.3	464	1	YSV5_CAEEL	Q10011 caenorhabdi
8	33	73.3	535	1	SPKC_SYNY3	P74745 synechocyst
9	33	73.3	1385	1	FAT1_SCHPO	O13735 schizosacch
10	33	73.3	1416	1	BLM_MOUSE	O88700 mus musculu
11	32	71.1	527	1	YAD7_SCHPO	Q09833 schizosacch
12	32	71.1	532	1	ICP0_HSVEB	P28990 equine herp
13	32	71.1	553	1	Q002_MYCTU	Q10381 mycobacteri
14	32	71.1	2261	1	ABC1_HUMAN	O95477 homo sapien
15	32	71.1	2261	1	ABC1_MOUSE	P41233 mus musculu
16	32	71.1	3801	1	LYST_HUMAN	Q99698 homo sapien
17	31	68.9	98	1	Y528_METJA	Q57948 methanococc
18	31	68.9	255	1	HCD2_DROME	O18404 drosophila
19	31	68.9	333	1	BIOB_HAEIN	P44987 haemophilus
20	31	68.9	366	1	GHSP_PIG	Q95254 sus scrofa
21	31	68.9	427	1	MFGM_RAT	P78490 rattus norv
22	31	68.9	463	1	MFGM_MOUSE	P21956 mus musculu
23	31	68.9	665	1	SIN1_SCHPO	Q9P7Y9 schizosacch
24	31	68.9	1062	1	UL78_HCMVA	P17149 human cytom
25	30	66.7	64	1	YC33_SKIFO	Q96888 skeletonema
26	30	66.7	81	1	DLTC_LACCA	P55153 lactobacill
27	30	66.7	195	1	RSEA_HAEIN	P44791 haemophilus
28	30	66.7	265	1	MOPA_RHOCA	Q08385 rhodobacter
29	30	66.7	311	1	PYFD_BACSU	P25906 bacillus su
30	30	66.7	322	1	PM18_YEAST	P36520 saccharomyc
31	30	66.7	389	1	YPP6_CAEEL	P41884 caenorhabdi
32	30	66.7	396	1	Y078_METJA	Q57972 methanococc
33	30	66.7	418	1	EPF1_RALSO	P58595 ralstonia s
34	30	66.7	441	1	GAT6_RAT	P46153 rattus norv
35	30	66.7	444	1	GAT6_MOUSE	Q61169 mus musculu
36	30	66.7	449	1	GAT6_HUMAN	Q92908 homo sapien
37	30	66.7	485	1	Y136_TREPA	O83172 treponema p
38	30	66.7	492	1	GATB_ARCFU	O28164 archaeoglob
39	30	66.7	504	1	Y104_MYCTH	Q10898 mycobacteri
40	30	66.7	516	1	YIJ7_YEAST	P40492 saccharomyc
41	30	66.7	535	1	VP11_VACCP	P08893 vaccinia vi
42	30	66.7	535	1	VP11_VACCO	P08894 vaccinia vi
43	30	66.7	635	1	VP12_VARV	P33871 variola vir
44	30	66.7	652	1	YD84_HUMAN	Q9P2Q3 homo sapien
45	30	66.7	703	1	PALY_BROFI	Q42609 bromheadia

ALIGNMENTS

RESULT 1
MFGM_HUMAN
ID MFGM_HUMAN STANDARD; PRT; 387 AA.
AC Q08431,
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMFG)
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
GN MFGE8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast, and Breast carcinoma;
RX MEDLINE=96214008; PubMed=8626764;
RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
RT "Cloning and sequence analysis of human breast epithelial antigen
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal
RT growth factor-like domain.";
RL DNA Cell Biol. 15:281-286(1996).
RN [2]
RP SEQUENCE OF 170-387 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=91371351; PubMed=1909932;
RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
RA Ceriani R.L.;
RT "A Mr 46,000 human milk fat globule protein that is highly expressed
RT in human breast tumors contains factor VIII-like domains.";
RL Cancer Res. 51:4994-4998(1991).
RN [3]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=98194924; PubMed=9537276;
RA Giuffrida M.G., Cavaletto M., Giunta C., Conti A.,
RA Godovac-Zimmermann J.;
RT "Isolation and characterization of full and truncated forms of human
RT breast carcinoma protein BA46 from human milk fat globule membranes.";
RL J. Protein Chem. 17:143-148(1998).
RN [4]
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
RX MEDLINE=99342078; PubMed=10411933;
RA Haeggqvist B., Naeslund J., Slotten K., Westermark G.T., Mucchiano G.,
RA Tjernberg L.O., Nordstedt C., Engstroem U., Westermark P.;
RT "Medin: an integral fragment of aortic smooth muscle cell-produced
RT lactadherin forms the most common human amyloid.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:8864-8869(1999).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=94105885; PubMed=8289323;
RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
RT (RGD)-dependent cell adhesion.";

```
RL   DNA Cell Biol 16:861-869(1997)
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
CC       SPECIFICALLY TO B-FLAVIVIRUS AND INHIBITS ITS REPLICATION.
CC   -!- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
CC       MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC   -!- PTM: MEDIN HAS A BLOCKED N-TERMINUS WITH MINOR SPECIES STARTING AT
CC       AMINO ACID 264 AND 273.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U58516; AAC50549.1; -.
DR   EMBL; S56151; AAB19771.1; -.
DR   MIM; 600328; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid
FT   SIGNAL       1     23
FT   CHAIN       24    387       LACTADHERIN.
FT   CHAIN      202    387       LACTADHERIN; SHORT FORM
FT   CHAIN      268    317       MEDIN.
FT   DOMAIN      24     67       EGF-LIKE.
FT   DOMAIN      70    225       F5/8 TYPE C 1.
FT   DOMAIN     230    387       F5/8 TYPE C 2.
FT   SITE        46     48       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    27     38       BY SIMILARITY.
FT   DISULFID    32     55       BY SIMILARITY.
FT   DISULFID    57     66       BY SIMILARITY.
FT   DISULFID    70    225       BY SIMILARITY.
FT   DISULFID   212    216       BY SIMILARITY.
FT   DISULFID   230    387       BY SIMILARITY
FT   CARBOHYD   238    238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   325    325       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   329    329       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   350    350       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   387 AA,  43123 MW,  2FF6571D-EC837B2D CRC64;


  Query Match             100.0%;  Score 45;  DB 1;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.062;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
           |||||||||
Db     356 NLFETPILA 364


RESULT   2
NU5M_GADMO
ID   NU5M_GADMO     STANDARD;      PRT;   612 AA.
AC   P55782;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3)
GN   MTND5 OR ND5.
OS   Gadus morhua (Atlantic cod).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidei; Gadidae;
OC   Gadus.
OX   NCBI_TaxID=8049;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NORWEGIAN COASTAL 1;
RX   MEDLINE=98141425; PubMed=9817926;
RA   Johansen S., Bakke I.;
RT   "The complete mitochondrial DNA sequence of Atlantic cod (Gadus
RT   morhua): relevance to taxonomic studies among codfishes.";
RL   Mol. Mar. Biol. Biotechnol. 5:203-214(1996).
CC   -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X99772; CAA68115.1; -.
DR   InterPro; IPR003916; NADHdh_oxdrdase5.
DR   InterPro; IPR001750; Oxidored_q1.
DR   InterPro; IPR001516; Oxidored_q1_N
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
DR   PRINTS; PR01434; NADHDHGNASE5.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ   SEQUENCE   612 AA,  68604 MW;  4AB69E2F1EDA150 CRC64;


  Query Match             80.0%;  Score 45;  DB 1;  Length 612;
  Best Local Similarity   75.0%;  Pred. No. 7.7;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPIL 8
             |||:||:|
Db     479 NLFKTPVL 486


RESULT   3
MFGM_PIG
ID   MFGM_PIG       STANDARD;      PRT;   409 AA.
AC   P79385;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Lactadherin (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM) (Sperm
DE   surface protein SP47) (PP47).
GN   MFGE8.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC       SPERMATOZOAN.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; Y11683; CAA72379.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Glycoprotein; Repeat; EGF-like domain.
FT   DOMAIN          2     41       EGF-LIKE 1.
FT   DOMAIN         44     88       EGF-LIKE 2.
FT   DOMAIN         91    247       F5/8 TYPE C 1.
FT   DOMAIN        252    409       F5/8 TYPE C 2.
FT   SITE           67     69       CELL ATTACHMENT SITE (POTENTIAL)
FT   DISULFID        6     17       BY SIMILARITY.
FT   DISULFID       11     29       BY SIMILARITY.
FT   DISULFID       31     40       BY SIMILARITY.
FT   DISULFID       91    247       BY SIMILARITY.
FT   DISULFID      234    238       BY SIMILARITY.
FT   DISULFID      252    409       BY SIMILARITY.
FT   CARBOHYD       41     41       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      372    372       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   409 AA;  45725 MW;  B0C07AF80029027A CRC64;


  Query Match             77.8%;  Score 35;  DB 1;  Length 409;
  Best Local Similarity   75.0%;  Pred. No. 8.2;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0,  Gaps    0,

Qy       1 NLFETPIL 8
           |:|||| |
Db     378 NMFETPFL 385


RESULT   4
NRX4_DROME
ID   NRX4_DROME     STANDARD;      PRT;  1284 AA.
AC   Q94887; Q9VIU5;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neurexin IV precursor
GN   NRX OR CG6827.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Canton-S;
RX   MEDLINE=97133213; PubMed 8978610;
RA   Baumgartner S.W., Littleton J.T., Broadie K., Bhat M.A., Harbecke R.,
RA   Lengyel J.A., Chiquet-Ehrismann R., Prokop A., Bellen H.J.;
RT   "A Drosophila neurexin is required for septate junction and blood-
RT   nerve barrier formation and function.",
RL   Cell 87:1059-1068(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   INTERACTION WITH CORACLE.
RX   MEDLINE 98177172; PubMed 9508778;
RA   Ward R.E. IV, Lamb R.S., Fehon R.G.;
RT   "A conserved functional domain of Drosophila coracle is required for
RT   localization at the septate junction and has membrane-organizing
RT   activity.";
RL   J. Cell Biol. 140:1463-1473(1998).
RN   [4]
RP   INTERACTION WITH DISCS LOST.
RX   MEDLINE 99220394; PubMed=10102271;
RA   Bhat M.A., Izaddoost S., Lu Y., Cho K.O., Choi K.W., Bellen H.J.;
RT   "Discs lost, a novel multi PDZ domain protein, establishes and
RT   maintains epithelial polarity.";
RL   Cell 96:833-845(1999).
CC   -!- FUNCTION: SEEMS TO PLAY A ROLE IN THE FORMATION AND FUNCTION OF
CC       SEPTATE JUNCTIONS. IS REQUIRED FOR THE BLOOD-BRAIN BARRIER
CC       FORMATION.
CC   -!- SUBUNIT: THE C-TERMINAL REGION INTERACTS WITH CORACLE. INTERACTS
CC       WITH DISCS LOST IN CIS FORM.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- TISSUE SPECIFICITY: FOUND IN SEPTATE JUNCTIONS OF EPITHELIAL AND
CC       GLIAL CELLS.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 LAMININ G-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE NEUREXIN FAMILY.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
```

```
DR   EMBL; X86685; CAA60383.1; -.
DR   EMBL; AE003542; AAF49951.1; -.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0013962; Nrx.
DR   InterPro; IPR003585; 4.1m.
DR   InterPro; IPR000561; EGF-like
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 1.
DR   Pfam; PF00054; laminin_G; 2.
DR   SMART; SM00294; 4.1m; 1.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 1.
DR   SMART; SM00282; LamG; 4.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS01285; FA58C_1; 1.
DR   PROSITE; PS01286; FA58C_2; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 4.
KW   Glycoprotein; Cell adhesion; Signal; Transmembrane; Repeat.
FT   SIGNAL        1     35       POTENTIAL.
FT   CHAIN        36   1284       NEUREXIN IV.
FT   DOMAIN       36   1217       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1218   1238       POTENTIAL.
FT   DOMAIN     1239   1284       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       47    185       F5/8 TYPE C.
FT   DOMAIN      220    369       LAMININ G LIKE 1.
FT   DOMAIN      403    540       LAMININ G-LIKE 2.
FT   DOMAIN      542    579       EGF-LIKE 1.
FT   DOMAIN      824    962       LAMININ G LIKE 3.
FT   DOMAIN      962    999       EGF-LIKE 2.
FT   DOMAIN     1032   1183       LAMININ G-LIKE 4.
FT   CONFLICT      1      4       MRPP -> MSA (IN REF. 1).
FT   CONFLICT    926    926       E -> D (IN REF. 1).
SQ   SEQUENCE   1284 AA;  145467 MW;  9372C71AC70B3056 CRC64;


  Query Match             77.8%;  Score 35;  DB 1;  Length 1284;
  Best Local Similarity   66.7%;  Pred. No. 27;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
           |:|| ||:|
Db     154 NVFEVPTIA 162


RESULT   5
DUR1_YEAST
ID   DUR1_YEAST     STANDARD;      PRT:  1835 AA.
AC   P32528;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Urea amidolyase [Includes: Urea carboxylase (EC 6.3.4.6); Allophanate
DE   hydrolase (EC 3.5.1.54)].
GN   DUR1,2 OR YBR208C OR YBR1448.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92199240; PubMed=1902034;
RA   Genbauffe F.S., Cooper T.G.;
RT   "The urea amidolyase (DUR1,2) gene of Saccharomyces cerevisiae.";
RL   DNA Seq. 2:19-32(1991).
RN   [2]
RP   SEQUENCE OF 1-893 FROM N.A.
RC   STRAIN=S288C;
RA   Rieger M.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 873-1835 FROM N.A.
RC   STRAIN=S288C;
RA   Feldmann H., Mannhaupt G., Schwarzlose C., Vetter I.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1487-1835 FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=93377417; PubMed=8368014;
RA   Bussereau F., Mallet L., Gaillon L., Jacquet M.;
RT   "A 12.8 kb segment, on the right arm of chromosome II from
RT   Saccharomyces cerevisiae including part of the DUR1,2 gene, contains
RT   five putative new genes.";
RL   Yeast 9:797-806(1993).
CC   -!- FUNCTION: HYDROLYSIS OF UREA TO AMMONIA AND CO2.
CC   -!- CATALYTIC ACTIVITY: ATP + urea + CO(2) = ADP + phosphate + urea-1-
CC       carboxylate.
CC   -!- CATALYTIC ACTIVITY: Urea-1-carboxylate + H(2)O = 2 CO(2) + 2
CC       NH(3).
CC   -!- COFACTOR: BIOTIN.
CC    !  PATHWAY: ALLANTOIN AND ARGININE METABOLISM.
CC   -!- SUBUNIT: MONOMER.
CC   -!- INDUCTION: BY ALLOPHANATE OR ITS NON METABOLIZED ANALOG
CC       OXALURATE. REPRESSED IN THE PRESENCE OF READILY USED NITROGEN
CC       SOURCES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M64926; AAC41643.1; -.
DR   EMBL; Z36077; CAA85172.1; -.
DR   EMBL; U21187; CAA79695.1; -.
DR   PIR; S46082; S46082.
DR   HSSP; P24182; 1BNC.
DR   SGD; S0000412; DUR1,2.
DR   InterPro; IPR000120; Amidase.
DR   InterPro; IPR001882; Biotin.
DR   InterPro; IPR000089; Biotin_lipoyl.
DR   InterPro; IPR000901; CPSase.
DR   InterPro; IPR003778; DUF183.
DR   InterPro; IPR003833; DUF213.
DR   Pfam; PF01425; Amidase; 1.
DR   Pfam; PF02785; Biotin_carb_C; 1.
DR   Pfam; PF00364; biotin_lipoyl; 1.
DR   Pfam; PF00289; CPSase_L_chain; 1.
DR   Pfam; PF02786; CPSase_L_D2; 1.
DR   Pfam; PF02626; DUF183; 1.
DR   Pfam; PF02682; DUF213; 1.
DR   PROSITE; PS00188; BIOTIN; 1.
DR   PROSITE; PS00866; CPSASE_1; 1.
DR   PROSITE; PS00867; CPSASE_2; 1.
KW   Ligase; Hydrolase; Multifunctional enzyme; Arginine metabolism;
KW   Biotin; ATP-binding.
FT   NP_BIND     122    129       ATP (POTENTIAL).
FT   BINDING    1798   1798       BIOTIN.
FT   SIMILAR    1779   1835       TO OTHER BIOTIN CARBOXYLASES C-TERMINI.
FT   CONFLICT     96     96       P -> R (IN REF. 1).
FT   CONFLICT    256    258       LKK -> KKN (IN REF. 1).
FT   CONFLICT    459    459       I -> M (IN REF. 1).
FT   CONFLICT    830    830       E -> K (IN REF. 1).
FT   CONFLICT   1395   1395       D -> E (IN REF. 1).
SQ   SEQUENCE   1835 AA;  201831 MW;  F52B0D60FE42CD65 CRC64;


  Query Match             77.8%;  Score 35;  DB 1;  Length 1835;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 FETPILA 9
```

```
          |||||||
Db    1418 FETPILA 1424


RESULT   6
MFGM_BOVIN
ID   MFGM_BOVIN     STANDARD;      PRT;   427 AA.
AC   Q95114; Q27959; P79344;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
DE   (MGP57/53) (PAS 6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
DE   SP47) (BP47) (Components 15/16).
GN   MFGE8.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC   STRAIN=HOLSTEIN; TISSUE=Mammary gland;
RX   MEDLINE=97008954; PubMed=8856064;
RA   Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RA   Petersen T.E.;
RT   "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT   milk fat globules.";
RL   Eur. J. Biochem. 240:628-636(1996).
RN   [2]
RP   SEQUENCE OF 18-427 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=96125736; PubMed=8541316;
RA   Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA   Matsuda T.;
RT   "Molecular cloning of glycoprotein antigens MGP-1, -2 recognized by
RT   monoclonal antibodies raised against bovine milk fat globule
RT   membrane.";
RL   Biochim. Biophys. Acta 1245:385-391(1995).
RN   [3]
RP   SEQUENCE OF 19-427 FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 116-140, 174-187, 253-276 AND 422-427.
RC   TISSUE=Milk;
RX   MEDLINE=93259574; PubMed=8485470;
RA   Mather I.H., Banghart L.R., Lane W.S.;
RT   "The major fat-globule membrane proteins, bovine components 15/16 and
RT   guinea-pig GP 55, are homologous to MFG-E8, a murine glycoprotein
RT   containing epidermal growth factor-like and factor V/VIII-like
RT   sequences.";
RL   Biochem. Mol. Biol. Int. 29:545-554(1993).
CC   -!- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
CC       MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM
CC       LACKS 53 AMINO ACIDS WITHIN THE F5/8 TYPE C 1 DOMAIN.
CC   -!- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
CC   -!- PTM: THE 2 O-LINKED GLYCANS CONSIST OF GAL, GLCNAC AND FUC, WITH
CC       PROBABLY FUC AS REDUCING TERMINAL SUGAR.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X91895; CAA62997.1; -.
DR   EMBL; S80643; AAB35894.2; -.
DR   EMBL; Y11719; CAA72406.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain;
KW   Alternative splicing.
FT   SIGNAL       1     18
FT   CHAIN       19    427       LACTADHERIN.
FT   DOMAIN      20     59       EGF-LIKE 1.
FT   DOMAIN      62    106       EGF-LIKE 2.
FT   DOMAIN     169    265       F5/8 TYPE C 1.
FT   DOMAIN     270    427       F5/8 TYPE C 2.
FT   SITE        85     87       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    24     35       BY SIMILARITY.
FT   DISULFID    29     47       BY SIMILARITY.
FT   DISULFID    49     58       BY SIMILARITY.
FT   DISULFID    66     77       BY SIMILARITY.
FT   DISULFID    71     94       BY SIMILARITY.
FT   DISULFID    96    105       BY SIMILARITY.
FT   DISULFID   109    265
FT   DISULFID   252    256
FT   DISULFID   270    427
FT   CARBOHYD    27     27       O-LINKED (FUC...) (IN PAS-6).
FT   CARBOHYD    34     34       O-LINKED (FUC...) (IN PAS-7).
FT   CARBOHYD    56     56       N-LINKED (GLCNAC...) (HYBRID) (IN PAS-6
FT                               AND PAS-7).
FT   CARBOHYD   227    227       N-LINKED (GLCNAC...) (HIGH MANNOSE) (IN
FT                               PAS-6).
FT   VARSPLIC   169    221       MISSING (IN SHORT ISOFORM).
FT   CONFLICT    19     19       A -> F (IN REF. 1).
FT   CONFLICT    28     28       L -> Q (IN REF. 1).
SQ   SEQUENCE   427 AA;  47411 MW;  4CBBEE3A1DC4EB24 CRC64;


  Query Match          73.3%;  Score 33;  DB 1;  Length 427;
  Best Local Similarity   66.7%;  Pred. No. 22;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NIFFTPILA 9
           |:|||| |
Db     396 NIFETPFQA 404


RESULT   7
YSV5_CAEEL
ID   YSV5_CAEEL     STANDARD;      PRT;   464 AA.
AC   Q10011;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 51.9 kDa protein T19C3.5 in chromosome III precursor.
GN   T19C3.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Fulton L.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: SOME IG-LIKE/FNIII/EGF/PLAC FAMILY.
```

```
CC   ---------- ----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; U28412; AAC46596.1; -.
DR   WormPep; T19C3.5; CE02096.
DR   InterPro; IPR001124; LBP_BPI_CETP.
DR   Pfam; PF01273; LBP_BPI_CETP; 1.
DR   Pfam; PF02886; LBP_BPI_CETP_C; 1.
DR   SMART; SM00328; BPI1; 1.
DR   SMART; SM00329; BPI2; 1.
DR   PROSITE; PS00400; LBP_BPI_CETP; FALSE_NEG.
KW   Hypothetical protein; Signal.
FT   SIGNAL        1     17       POTENTIAL
FT   CHAIN        18    464       HYPOTHETICAL PROTEIN T19C3.5.
SQ   SEQUENCE   464 AA;  51939 MW;  3029CFD72F65E67F CRC64;


  Query Match             73.4%;  Score 33;  DB 1;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETP 6
             ||||||
Db     134 NLFETP 139


RESULT   8
SPKC_SYNY3
ID   SPKC_SYNY3     STANDARD;      PRT;   535 AA.
AC   P74745;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable serine/threonine-protein kinase C (EC 2.7.1.-).
GN   SPKC OR SLR0599.
OS   Synechocystis sp. (strain PCC 6803)
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kamei A., Ikeuchi M.;
RT   "A novel gene, spkC, encodes active Ser/Thr protein kinase in the
RT   motile cyanobacterium Synechocystis sp. PCC 6803.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   ----------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; AB046599; BAB17035.1; -.
DR   EMBL; D90917; BAA18865.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Complete proteome.
FT   DOMAIN       12    277       PROTEIN KINASE.
FT   NP_BIND      18     26       ATP (BY SIMILARITY).
FT   BINDING      43     43       ATP (BY SIMILARITY).
FT   ACT_SITE    142    142       BY SIMILARITY.
SQ   SEQUENCE   535 AA;  58141 MW;  533CD9FC0E3D24D8 CRC64;


  Query Match             73.3%;  Score 33;  DB 1;  Length 535;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETP 6
             ||||||
Db     396 NLFETP 401


RESULT   9
FAT1_SCHPO
ID   FAT1_SCHPO     STANDARD;      PRT;  1385 AA.
AC   O13735; Q9UTJ0;
DT   15-JUL-1998 (Rel. 36, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Actin interacting protein 3 homolog.
GN   FAT1 OR SPAC15A10.15 OR SPAC15F1.01.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE OF 1-1033 FROM N.A.
RC   STRAIN=972;
RA   Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1023-1385 FROM N.A.
RC   STRAIN=972;
RA   Murphy L., Harris D., Barrell B.G., Rajandream M.A., Lyne M.H.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   IDENTIFICATION, AND GENE NAME.
RX   MEDLINE=20143585; PubMed=10679021;
RA   Jin H., Amberg D.C.;
RT   "The secretory pathway mediates localization of the cell polarity
RT   regulator Aip3p/Rud6p.";
RL   Mol. Biol. Cell 11:647-661(2000).
CC   -!- FUNCTION: INVOLVED IN THE ORGANIZATION AND/OR FUNCTION OF THE
CC       ACTIN CYTOSKELETON.
CC   -!- SIMILARITY: TO YFAST BUD6
CC   ----------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; Z97208; CAB10112.1; -.
DR   EMBL; AL109772; CAB61210.1; -.
KW   Coiled coil; Cytoskeleton.
FT   DOMAIN     1009   1096       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1385 AA;  154325 MW;  ABB3D40CC4FF7537 CRC64;
```

```
  Query Match             73.3%;  Score 33;  DB 1;  Length 1385;
  Best Local Similarity   62.5%;  Pred. No. 77;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0,  Gaps    0;

Qy       1 NLPETPIL 8
           |::||||:
Db    1304 NIYPTPIV 1311


RESULT  10
BLM_MOUSE
ID   BLM_MOUSE      STANDARD;      PRT;  1416 AA.
AC   O88700; O88198;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Bloom's syndrome protein homolog (EC 3.6.1.-) (mBLM).
GN   BLM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND MUTAGENESIS.
RX   MEDLINE=99054654; PubMed=9840919;
RA   Bahr A., de Graeve F., Kedinger C., Chatton B.;
RT   "Point mutations causing Bloom's syndrome abolish ATPase and DNA
RT   helicase activities of the BLM protein.";
RL   Oncogene 17:2565-2571(1998).
RN   [2]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   STRAIN=BALB/C; TISSUE=Testis, Spermatocyte, and Brain;
RX   MEDLINE=98322127; PubMed=9655940;
RA   Seki T., Wang W.-S., Okumura N., Seki M., Katada T., Enomoto T.;
RT   "cDNA cloning of mouse BLM gene, the homologue to human Bloom's
RT   syndrome gene, which is highly expressed in the testis at the mRNA
RT   level.";
RL   Biochim. Biophys. Acta 1398:377-381(1998).
CC   -!- FUNCTION: PARTICIPATES IN DNA REPLICATION AND MAY PARTICIPATE IN
CC       REPAIR. EXHIBITS A MAGNESIUM-DEPENDENT ATP-DEPENDENT DNA-HELICASE
CC       ACTIVITY THAT UNWINDS SINGLE- AND DOUBLE-STRANDED DNA IN A 3'-5'
CC       DIRECTION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS 12-14 DAYS AFTER
CC       BIRTH (CORRESPONDING TO THE PACHYTENE PHASE) AND AT MUCH LOWER
CC       LEVELS IN BRAIN, HEART, KIVER, LUNG, THYMUS, KIDNEY AND SPLEEN.
CC   -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RECQ SUBFAMILY
CC   -!- SIMILARITY: CONTAINS 1 HRDC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z98263; CAB10933.1; -.
DR   EMBL; AB008674; BAA32001.1; -.
DR   MGD; MGI:1328362; Blm.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002464; DEAH_ATP_helicase.
DR   InterPro; IPR002121; HRDC.
DR   InterPro; IPR000330; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00570; HRDC; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00341; HRDC; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; 1.
KW   Hydrolase, Helicase, ATP-binding; DNA-binding; Nuclear protein;
KW   DNA replication.
FT   DOMAIN      219    231       POLY-GLU.
FT   DOMAIN      564    574       POLY-ASP.
FT   DOMAIN     1312   1317       POLY-GLU.
FT   NP_BIND     697    704       ATP (BY SIMILARITY).
FT   SITE        803    806       DEAH BOX.
FT   DOMAIN     1217   1297       HRDC.
FT   DOMAIN     1333   1348       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   MUTAGEN     680    680       Q->P: REDUCED ATPASE AND HELICASE
FT                                ACTIVITIES.
FT   MUTAGEN     703    703       K->A: REDUCED ATPASE AND HELICASE
FT                                ACTIVITIES.
FT   MUTAGEN     849    849       I->T: REDUCED ATPASE AND HELICASE
FT                                ACTIVITIES.
FT   MUTAGEN    1063   1063       C->S: REDUCED ATPASE AND HELICASE
FT                                ACTIVITIES.
FT   CONFLICT    131    131       L -> P (IN REF. 2).
FT   CONFLICT    229    229       E -> EE (IN REF. 2).
FT   CONFLICT    535    535       V -> M (IN REF. 2).
FT   CONFLICT    546    547       WN -> RT (IN REF. 2).
FT   CONFLICT    574    574       MISSING (IN REF. 2).
FT   CONFLICT    591    591       T -> A (IN REF. 2).
FT   CONFLICT    621    621       I -> N (IN REF. 2).
FT   CONFLICT   1295   1295       V -> L (IN REF. 2).
SQ   SEQUENCE   1416 AA;  158365 MW;  447C8110A775DD42 CRC64;


  Query Match             73.3%;  Score 33;  DB 1;  Length 1416;
  Best Local Similarity   75.0%;  Pred. No. 78;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0,  Gaps    0;

Qy       1 NLPETPIL 8
           |||| |:|
Db     490 NLPERPLL 497


RESULT  11
YAD7_SCHPO
ID   YAD7_SCHPO     STANDARD;      PRT;   527 AA.
AC   Q09833;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 59.6 kDa protein C4G8.07C in chromosome I.
GN   SPAC4G8.07C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Badcock K., Churcher C.M., Borrell B.J., Rajandream M.A., Walsh S.V.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS EXONUCLEASE RUVC (RNC1).
CC   -!- SIMILARITY: SOME, TO THE RNA METHYLTRANSFERASE TRMA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z56276; CAA91208.1; -.
DR   InterPro; IPR002792; DUF90.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR001566; TRMA_1.
DR   Pfam; PF01938; TRAM; 1.
DR   PROSITE; PS01230; TRMA_1; 1.
```

```
DR   PROSITE; PS01231; TRMA_2; 1.
KW   Hypothetical protein; Hydrolase; Nuclease; Transferase;
KW   Methyltransferase.
FT   ACT_SITE     479     479       BY SIMILARITY.
SQ   SEQUENCE   527 AA;  59613 MW;  ACCE80D8C6FFD77F CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 527;
  Best Local Similarity   66.7%;  Pred. No. 45;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
             || | |:||
Db      96 NLIEKPVLA 104


RESULT  12
ICP0_HSVEB
ID   ICP0_HSVEB     STANDARD;      PRT;   532 AA.
AC   P28990;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Trans-acting transcriptional protein ICP0.
GN   63.
OS   Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=31520;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92295566; PubMed=1318606;
RA   Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT   "The DNA sequence of equine herpesvirus-1.";
RL   Virology 189:304-316(1992).
RN   [2]
RP   STRUCTURE BY NMR OF 1-63.
RX   MEDLINE=94087718; PubMed=8263911;
RA   Everett R.D., Barlow P.N., Milner A., Luisi B., Orr A., Hope G.,
RA   Lyon D.;
RT   "A novel arrangement of zinc binding residues and secondary structure
RT   in the C3HC4 motif of an alpha herpes virus protein family.";
RL   J. Mol. Biol. 234:1038-1047(1993).
RN   [3]
RP   STRUCTURE BY NMR OF 1-63.
RX   MEDLINE=94172642; PubMed=8126734;
RA   Barlow P.N., Luisi B., Milner A., Elliott M., Everett R.D.;
RT   "Structure of the C3HC4 domain by 1H nuclear magnetic resonance
RT   spectroscopy. A new structural class of zinc-finger.";
RL   J. Mol. Biol. 237:201-211(1994).
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M86664; AAB02498.1; -.
DR   PIR; I36801; W2BEF5.
DR   PDB; 1CHC; 30-APR-94.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
KW   DNA-binding; 3D-structure.
FT   ZN_FING       8     47       RING-TYPE.
FT   METAL         8      8       ZINC 1.
FT   METAL        11     11       ZINC 1.
FT   METAL        24     24       ZINC 2.
FT   METAL        26     26       ZINC 2.
FT   METAL        29     29       ZINC 1.
FT   METAL        32     32       ZINC 1.
FT   METAL        43     43       ZINC 2.
FT   METAL        46     46       ZINC 2.
FT   DOMAIN      210    217       POLY-SER.
FT   STRAND       19     21
FT   TURN         22     25
FT   STRAND       26     28
FT   TURN         30     31
FT   HELIX        32     39
FT   TURN         43     46
FT   STRAND       53     54
SQ   SEQUENCE   532 AA;  58629 MW;  B4CB7E16FA26FDFA CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 532;
  Best Local Similarity   87.5%;  Pred. No. 46;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LFETPILA 9
             ||||| ||
Db     136 LFETPELA 143


RESULT  13
ODO2_MYCTU
ID   ODO2_MYCTU     STANDARD;      PRT;   553 AA.
AC   O10381;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Dihydrolipoamide succinyltransferase component of 2-oxoglutarate
DE   dehydrogenase complex (EC 2.3.1.61) (E2).
GN   SUCB OR PV2215 OR MT2272 OR MTCY190.26.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN="Erdman 1551 / Oshkosh";
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 2-OXOGLUTARATE DEHYDROGENASE COMPLEX CATALYZES THE
CC       OVERALL CONVERSION OF 2-OXOGLUTARATE TO SUCCINYL-COA & CO(2). IT
CC       CONTAINS MULTIPLE COPIES OF 3 ENZYMATIC COMPONENTS: 2-OXOGLUTARATE
CC       DEHYDROGENASE (E1), DIHYDROLIPOAMIDE SUCCINYLTRANSFERASE (E2) AND
CC       LIPOAMIDE DEHYDROGENASE (E3) (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Succinyl-CoA + dihydrolipoamide = CoA + S-
```

```
CC       succinyldihydrolipoamide.
CC   -!- COFACTOR: CONTAINS TWO COVALENTLY-BOUND LIPOYL COFACTOR
CC       (POTENTIAL).
CC   -!- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL
CC       SYMMETRY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 LIPOYL BINDING DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z70283; CAA94256.1; -.
DR   EMBL; AE007072; AAK46557.1; -.
DR   HSSP; P07016; 1E2O.
DR   TIGR; MT2272; -.
DR   TubercuList; Rv2215; -.
DR   InterPro; IPR001078; 2Oxoacid_dh.
DR   InterPro; IPR000089; Biotin_lipoyl.
DR   InterPro; IPR003016; Lipoyl.
DR   InterPro; IPR004167; e3_binding.
DR   Pfam; PF00198; 2-oxoacid_dh; 1.
DR   Pfam; PF00364; biotin_lipoyl; 2.
DR   Pfam; PF02817; e3_binding; 1.
DR   ProDom; PD001115; 2Oxoacid_dh; 1.
DR   PROSITE; PS00189; LIPOYL; 2.
KW   Glycolysis; Transferase; Acyltransferase; Lipoyl; Complete proteome.
FT   BINDING      43     43       LIPOYL (POTENTIAL).
FT   BINDING     162    162       LIPOYL (POTENTIAL).
FT   ACT_SITE    523    523       BY SIMILARITY.
FT   ACT_SITE    527    527       BY SIMILARITY.
SQ   SEQUENCE   553 AA;  57087 MW;  54B6E70D23B804A7 CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 553;
  Best Local Similarity   85.7%;  Pred. No. 48;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LPETPIL 8
           |:||||
Db     474 LPDTPIL 480


RESULT  14
ABC1_HUMAN
ID   ABC1_HUMAN     STANDARD;      PRT;  2261 AA.
AC   O95477; Q9UN08; Q9UN07; Q9UN06; Q9NQV4; Q9UN09; Q96T85; Q96S56;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette
DE   transporter 1) (ATP-binding cassette 1) (ABC-1) (Cholesterol efflux
DE   regulatory protein).
GN   ABCA1 OR ABC1 OR CERP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20345099; PubMed=10884428;
RA   Santamarina-Fojo S., Peterson K.M., Knapper C.L., Qiu Y.,
RA   Freeman L.A., Cheng J.-F., Osorio J., Remaley A.T., Yang X.-P.,
RA   Haudenschild C.C., Prades J., Chimini G., Blackmon F.E.,
RA   Francois T.L., Duverger N., Rubin E.M., Rosier M., Denefle P.,
RA   Fredrickson D.S., Brewer H.B. Jr.;
RT   "Complete genomic sequence of the human ABCA1 gene: analysis of the
RT   human and mouse ATP binding cassette A promoter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:7987-7992(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Schwartz K., Lawn R.M., Wade D.P.;
RT   "ABCA1 gene expression and apoA-I mediated cholesterol efflux are
RT   regulated by LXR.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21251004; PubMed=11352567;
RA   Qin Y., Cavelier L., Chiu S., Yang X., Rubin E., Cheng J.-F.;
RT   "Human and mouse ABCA1 comparative sequencing and transgenesis
RT   studies revealing novel regulatory sequences.";
RL   Genomics 73:65-76(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Tanaka A.R., Abe-Dohmae S., Arakawa R., Ohnishi T., Kitano A.,
RA   Kioka N., Amachi T., Yokoyama S., Ueda K.;
RT   "A new topological model of functional human ABCA1-signal peptide
RT   cleavage and glycosylation at a large extracellular domain.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 21-2261 FROM N.A.
RX   MEDLINE=99194549; PubMed=10092505;
RA   Langmann T., Klucken J., Reil M., Liebisch G., Luciani M.F.,
RA   Chimini G., Kaminski W.E., Schmitz G.;
RT   "Molecular cloning of the human ATP-binding cassette transporter 1
RT   (hABC1): evidence for sterol-dependent regulation in macrophages.";
RL   Biochem. Biophys. Res. Commun. 257:29-33(1999).
RN   [6]
RP   SEQUENCE OF 21-2261 FROM N.A.
RX   MEDLINE=99364413; PubMed=10431238;
RA   Rust S., Rosier M., Funke H., Real J., Amoura Z., Piette J.-C.,
RA   Deleuze J.-F., Brewer H.B., Duverger N., Denefle P., Assmann G.;
RT   "Tangier disease is caused by mutations in the gene encoding
RT   ATP-binding cassette transporter 1.";
RL   Nat. Genet. 22:352-355(1999).
RN   [7]
RP   VARIANTS FHA THR-1091 AND 1893-GLU-ASP-1894 DEL.
RX   MEDLINE=20001430; PubMed=10533863;
RA   Marcil M., Brooks-Wilson A., Clee S.M., Roomp K., Zhang L.-H., Yu L.,
RA   Collins J.A., van Dam M., Molhuizen H.O.F., Loubser O.,
RA   Ouelette B.F.F., Sensen C.W., Fichter K., Mott S., Denis M.,
RA   Boucher B., Pimstone S., Genest J. Jr., Kastelein J.J.P., Hayden M.R.;
RT   "Mutations in the ABC1 gene in familial HDL deficiency with defective
RT   cholesterol efflux.";
RL   Lancet 354:1341-1346(1999).
RN   [8]
RP   VARIANTS TD ARG-597 AND ARG-1477, AND VARIANT FHA LEU-693 DEL.
RX   MEDLINE=99364411; PubMed=10431236;
RA   Brooks-Wilson A., Marcil M., Clee S.M., Zhang L.-H., Roomp K.,
RA   van Dam M., Yu L., Brewer C., Collins J.A., Molhuizen H.O.F.,
RA   Loubser O., Ouelette B.F.F., Fichter K., Ashbourne-Excoffon K.J.D.,
RA   Sensen C.W., Scherer S., Mott S., Denis M., Martindale D.,
RA   Frohlich J., Morgan K., Koop B., Pimstone S., Kastelein J.J.P.,
RA   Hayden M.R.;
RT   "Mutations in ABC1 in Tangier disease and familial high-density
RT   lipoprotein deficiency.";
RL   Nat. Genet. 22:336-345(1999).
RN   [9]
RP   VARIANTS TD SER-590; SER-935 AND VAL-937, AND VARIANTS ALA-299 AND
RP   MET-883.
RX   MEDLINE=99364412; PubMed=10431237;
RA   Bodzioch M., Orso E., Klucken J., Langmann T., Boettcher A.,
RA   Diederich W., Drobnik W., Barlage S., Buechler C.,
RA   Porsch-Oezcueruemez M., Kaminski W.E., Hahmann H.W., Oette K.,
RA   Rothe G., Aslanidis C., Lackner K.J., Schmitz G.;
RT   "The gene encoding ATP-binding cassette transporter 1 is mutated in
RT   Tangier disease.";
RL   Nat. Genet. 22:347-351(1999).
RN   [10]
RP   VARIANTS TD ILE 929, ARG 597 AND ARG-1477, AND VARIANTS FHA LEU-693
RP   DEL; THR 1091; 1893-GLU-ASP-1894 DEL AND LEU-2150.
```

RX MEDLINE 20540002; PubMed 11086027;
RA Clee S.M., Kastelein J.J.P., van Dam M., Marcil M., Roomp K.,
RA Zwarts K.Y., Collins J.A., Roelants R., Tamasawa N., Stulc T.,
RA Suda T., Ceska R., Boucher B., Rondeau C., DeSouich C.,
RA Brooks Wilson A., Molhuizen H.O.F., Frohlich J., Genest J. Jr.,
RA Hayden M.R.;
RT "Age and residual cholesterol efflux affect HDL cholesterol levels and
RT coronary artery disease in ABCA1 heterozygotes.",
RL J. Clin. Invest. 106:1253-1259(2000).
RN [11]
RP VARIANTS TD ASN-1289 AND HIS-1800.
RX MEDLINE 20171564; PubMed 10706591;
RA Brousseau M.E., Schaefer E.J., Dupuis J., Eustace B.,
RA Van Eerdewegh P., Goldkamp A.L., Thurston L.M., FitzGerald M.G.,
RA Yasek McKenna D., O'Neill G., Eberhart G.P., Welfenbach B.,
RA Ordovas J.M., Freeman M.W., Brown R.H. Jr., Gu J.Z.;
RT "Novel mutations in the gene encoding ATP binding cassette 1 in four
RT tangier disease kindreds.";
RL J. Lipid Res. 41:433-441(2000).
RN [12]
RP VARIANT TD ASP 1046, VARIANT FHA CYS-220, AND VARIANTS LYS-219,
RP ILE-825, MET-883 AND LYS-1587.
RX MEDLINE-20496633; PubMed 10938021;
RA Wang J., Burnett J.R., Near S., Young K., Zinman B., Hanley A.J.G.,
RA Connelly P.W., Harris S.B., Hegele R.A.;
RT "Common and rare ABCA1 variants affecting plasma HDL cholesterol.";
RL Arterioscler. Thromb. Vasc. Biol. 20:1983-1989(2000).
RN [13]
RP VARIANT TD TRP-587, AND VARIANT LEU-2168.
RX MEDLINE-21157003; PubMed 11257261;
RA Bertolini S., Pisciotta L., Seri M., Cusano R., Cantafora A.,
RA Calabresi L., Franceschini C., Ravazzolo R., Calandra S.;
RT "A point mutation in ABC1 gene in a patient with severe premature
RT coronary heart disease and mild clinical phenotype of Tangier
RT disease.";
RL Atherosclerosis 154:599-605(2001).
RN [14]
RP VARIANTS LYS-219; MET-883 AND ASP-1172.
RX MEDLINE 21157003; PubMed 11257261;
RA Brousseau M.E., Bodzioch M., Schaefer E.J., Goldkamp A.L., Kielar D.,
RA Probst M., Ordovas J.M., Aslanidis C., Lackner K.J.,
RA Bloomfield Rubins H., Collins D., Robins S.J., Wilson P.W.F.,
RA Schmitz G.;
RT "Common variants in the gene encoding ATP binding cassette transporter
RT 1 in men with low HDL cholesterol levels and coronary heart disease.",
RL Atherosclerosis 154:607-611(2001).
RN [15]
RP VARIANT TD LEU-1506.
RX MEDLINE-21469429; PubMed 11476961;
RA Lapicka-Bodzioch K., Bodzioch M., Knuoli M., Kielar D., Probst M.,
RA Kiec B., Andrikovics H., Boettcher A., Hubacek J., Aslanidis C.,
RA Suttorp N., Schmitz G.;
RT "Homogeneous assay based on 52 primer sets to scan for mutations of
RT the ABCA1 gene and its application in genetic analysis of a new
RT patient with familial high-density lipoprotein deficiency syndrome.";
RL Biochim. Biophys. Acta 1537:42-48(2001).
RN [16]
RP VARIANTS TD ASN-1289 AND TRP-2081, AND VARIANT LYS 219.
RX MEDLINE 21469433; PubMed 11476965;
RA Huang W., Moriyama K., Koja T., Hua R., Ageta M., Kawabata S.,
RA Mawatari K., Imamura T., Eto T., Kawamura M., Teramoto T., Sasaki J.;
RT "Novel mutations in ABCA1 gene in Japanese patients with Tangier
RT disease and familial high density lipoprotein deficiency with
RT coronary heart disease.";
RL Biochim. Biophys. Acta 1537:71-78(2001).
RN [17]
RP VARIANTS LYS-219; ALA-399; MET-771; PRO-774; ASN-776; ILE-825;
RP MET-883; ASP-1172; LYS-1587 AND CYS-1731.
RA MEDLINE 21138374; PubMed 11258261;
RA Clee S.M., Zwinderman A.H., Engert J.C., Zwarts K.Y.,
RA Molhuizen H.O.F., Roomp K., Jukema J.W., van Wijland M., van Dam M.,
RA Hudson T.J., Brooks Wilson A., Genest J. Jr., Kastelein J.J.P.,
RA Hayden M.R.;
RT "Common genetic variation in ABCA1 is associated with altered
RT lipoprotein levels and a modified risk for coronary artery disease.";
RL Circulation 103:1198-1205(2001).
RN [18]
RP VARIANT TD THR 255, AND VARIANT ATHEROSCLEROSIS ASP-1611.
RX MEDLINE=21645894; PubMed=11785958;
RA Nishida Y., Hirano K., Tsukamoto K., Nagano M., Ikegami C., Roomp K.,
RA Ishibata M., Sakane N., Zhang Z., Tsujii K., Matsuyama A., Ohama T.,
RA Matsuura F., Ishigami M., Sakai N., Hiraoka H., Hattori H.,
RA Wellington C., Yoshida Y., Misuqi S., Hayden M.R., Eqashira T.,
RA Yamashita S., Matsuzawa Y.;
RT "Expression and functional analyses of novel mutations of ATP-binding
RT cassette transporter-1 in japanese patients with high-density
RT lipoprotein deficiency.";
RL Biochem. Biophys. Res. Commun. 290:713-721(2002).
CC -!- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION
CC TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL
CC TRANSPORT.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED, BUT MOST ABUNDANT IN
CC MACROPHAGES.
CC -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.
CC -!- DISEASE: DEFECTS IN ABCA1 ARE A CAUSE OF HIGH DENSITY LIPOPROTEIN
CC DEFICIENCY TYPE I (HDLD1), ALSO KNOWN AS TANGIER DISEASE (TD). TD
CC IS A RECESSIVE DISORDER CHARACTERIZED BY ABSENCE OF HIGH DENSITY
CC LIPOPROTEIN (HDL) CHOLESTEROL FROM PLASMA, HEPATOSPLENOMEGALY,
CC PERIPHERAL NEUROPATHY, AND FREQUENTLY PREMATURE CORONARY ARTERY
CC DISEASE (CAD).
CC -!- DISEASE: Defects in ABCA1 are a cause of high density lipoprotein

Query Match 71.1%; Score 42; DB 1; Length 2261;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPI 7
||||:|:
Db 808 NLFESPV 814

RESULT 15
ABC1_MOUSE
ID ABC1_MOUSE STANDARD; PRT; 2261 AA.
AC P41233;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family A, member 1 (ATP binding cassette
DE transporter 1) (ATP-binding cassette 1) (ABC-1).
GN ABCA1 OR ABC1
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DBA/2; TISSUE=Macrophage;
RX MEDLINE=94275009; PubMed=8099792;
RA Luciani M.F., Denizot F., Savary S., Mattei M.-G., Chimini G.;
RT "Cloning of two novel ABC transporters mapping on human chromosome
RT 9.";
RL Genomics 21:150-159(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Qiu Y., Cavelier L., Chiu S., Rubin E., Cheng J.-F.;
RT "Human and mouse ABCA1 comparative sequencing and transgenesis studies
RT identify potential regulatory sequences.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION
CC TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL
CC TRANSPORT (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED IN ADULT TISSUES. HIGHEST

```
CC       LEVELS ARE FOUND IN PREGNANT UTERUS AND UTERUS.
CC   -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC       EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC       ATP BINDING CASSETTE (ABC) DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X75926; CAA53530.1; ALT_INIT.
DR   EMBL; AF287263; AAG39073.1; ALT_INIT.
DR   MGD; MGI:99607; Abca1.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_CTP_A.
DR   InterPro; IPR003838; DUF214.
DR   InterPro; IPR000897; SRP54.
DR   Pfam; PF00005; ABC_tran; 2.
DR   Pfam; PF02687; DUF214; 1.
DR   Pfam; PF00448; SRP54; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Glycoprotein; Transmembrane; Transport.
FT   TRANSMEM     26     42       POTENTIAL.
FT   TRANSMEM    640    656       POTENTIAL.
FT   TRANSMEM    690    706       POTENTIAL.
FT   TRANSMEM    717    733       POTENTIAL.
FT   TRANSMEM    749    765       POTENTIAL.
FT   TRANSMEM    771    787       POTENTIAL.
FT   TRANSMEM   1041   1057       POTENTIAL.
FT   TRANSMEM   1351   1367       POTENTIAL.
FT   TRANSMEM   1661   1677       POTENTIAL.
FT   TRANSMEM   1708   1724       POTENTIAL.
FT   TRANSMEM   1737   1753       POTENTIAL.
FT   TRANSMEM   1775   1791       POTENTIAL.
FT   TRANSMEM   1854   1870       POTENTIAL.
FT   NP_BIND     933    940       ATP (POTENTIAL).
FT   NP_BIND    1946   1953       ATP (POTENTIAL).
FT   CARBOHYD     14     14       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    244    244       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    337    337       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    349    349       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400    400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    489    489       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    521    521       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    820    820       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1144   1144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1294   1294       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1453   1453       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1499   1499       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1504   1504       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1637   1637       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2044   2044       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2238   2238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1567   1568       MISSING (IN REF. 2).
FT   CONFLICT   2024   2024       MISSING (IN REF. 2).
SQ   SEQUENCE   2261 AA;  254011 MW;  FAE62B21FD1D09F9 CRC64;


  Query Match          71.1%;  Score 32;  DB 1;  Length 2261;
  Best Local Similarity  71.4%;  Pred. No. 2.1e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 NLFETPI 7
            ||||:|:
Db      808 NLFESPV 814


Search completed: September  5, 2002, 15:31:41
Job time: 476 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:01 ; Search time 122.86 Seconds
(without alignments)
12,673 Million cell updates/sec

Title: US-09-744-804-37
Perfect score: 45
Sequence: 1 NLFETPILA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	100.0	335	4	Q9BTL9	Q9btl9 homo sapien
2	39	86.7	363	6	O77718	O77718 equus cabal
3	36	80.0	541	17	Q972X2	Q972x2 sulfolobus
4	36	80.0	669	10	Q9LZ27	Q9lz27 arabidopsis
5	36	80.0	705	10	Q9LT29	Q9lt29 arabidopsis
6	34	75.6	250	16	O06544	O06544 mycobacteri
7	34	75.6	1244	2	Q9RAH3	Q9rah3 nostoc sp.
8	33	73.3	245	15	Q75814	Q75814 human immun
9	33	73.3	304	16	Q92AH5	Q92ah5 listeria in
10	33	73.3	572	5	O44508	O44508 caenorhabdi
11	33	73.3	658	3	O74798	O74798 schizosacch
12	33	73.3	694	5	Q21225	Q21225 caenorhabdi
13	33	73.3	1162	4	Q9BV54	Q9bv54 homo sapien
14	33	73.3	1241	4	Q9Y4B4	Q9y4b4 homo sapien
15	33	73.3	1466	11	Q99NG0	Q99ng0 mus musculu
16	33	73.3	2205	5	Q9NGQ2	Q9ngq2 dictyosteli
17	32	71.1	78	4	P78328	P78328 homo sapien
18	32	71.1	121	17	O58395	O58395 pyrococcus
19	32	71.1	173	16	Q9KT24	Q9kt24 vibrio chol
20	32	71.1	208	17	O59231	O59231 pyrococcus
21	32	71.1	245	16	Q9A1C1	Q9a1c1 streptococc
22	32	71.1	255	2	Q9AHY1	Q9ahy1 pseudomonas
23	32	71.1	255	16	Q91010	Q91010 pseudomonas
24	32	71.1	268	16	Q92CF2	Q92cf2 rhizobium m
25	32	71.1	278	16	Q9ABA4	Q9aba4 caulobacter
26	32	71.1	311	16	Q98HW3	Q98hw3 rhizobium l
27	32	71.1	328	15	Q9EC03	Q9ec03 human immun
28	32	71.1	328	15	Q9ECE3	Q9ece3 human immun
29	32	71.1	333	5	Q9VN80	Q9vn80 drosophila
30	32	71.1	394	16	Q9PED1	Q9ped1 xylella fas
31	32	71.1	398	16	Q9HW26	Q9hw26 pseudomonas
32	32	71.1	399	10	Q39254	Q39254 arabidopsis
33	32	71.1	399	16	Q9HUP4	Q9hup4 pseudomonas
34	32	71.1	417	3	Q9UTD0	Q9utd0 schizosacch
35	32	71.1	419	12	P87662	P87662 equine herp
36	32	71.1	425	16	Q9PKQ6	Q9pkq6 chlamydia m
37	32	71.1	433	16	Q9F1S7	Q9f1s7 bacillus ha
38	32	71.1	445	2	O52293	O52293 agrobacteri
39	32	71.1	456	10	Q9FF23	Q9ff23 arabidopsis
40	32	71.1	530	16	O52959	O52959 mycobacteri
41	32	71.1	586	1	Q9C4P0	Q9c4p0 halobacteri
42	32	71.1	590	2	Q57164	Q57164 acinetobact
43	32	71.1	670	10	Q9ZQM7	Q9zqm7 arabidopsis
44	32	71.1	779	3	Q12168	Q12168 saccharomyc
45	32	71.1	879	16	P71076	P71076 bacillus su

ALIGNMENTS

RESULT 1
Q9BTL9
ID Q9BTL9 PRELIMINARY; PRT; 335 AA.
AC Q9BTL9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO MILK FAT GLOBULE-EGF FACTOR 8 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA.;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
DR EMBL; BC003610; AAH03610.1; -.
DR HSSP; P08709; 1BF9.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR000421; FA58_C.
DR PRINTS; PR00010; EGFBLOOD.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00201; EGF_like; 1.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01285; FA58C_1; 1.
DR PROSITE; PS01286; FA58C_2; 2.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 335 AA, 37523 MW, 20E84E5BB78E8E25 CRC64;

Query Match 100.0%; Score 45; DB 4; Length 335;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NLFETPILA 9
           |||||||||
Db     404 NLFETPILA 412


RESULT   2
O77718
ID   O77718      PRELIMINARY;      PRT;   363 AA.
AC   O77718;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SPERM-MEMBRANE ASSOCIATED PROTEIN P47 (FRAGMENT).
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Gentzel M., Toepfer-Petersen E.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ010121; CAA09010.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58C
DR   Pfam; PF00008; EGF; 1
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 1.
KW   EGF-like domain; Glycoprotein.
FT   NON_TER       1       1
FT   NON_TER     363     363
SQ   SEQUENCE   363 AA;  40744 MW;  1F8B6395AF32338D CRC64;


  Query Match             86.7%;  Score 39;  DB 6;  Length 363;
  Best Local Similarity   77.8%;  Pred. No. 3.9;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
           |:|||| ||
Db     339 NMFETPFLA 347


RESULT   3
Q972X2
ID   Q972X2      PRELIMINARY;      PRT;   531 AA.
AC   Q972X2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL PROTEIN ST1017.
GN   ST1017.
OS   Sulfolobus tokodaii.
OC   Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX   NCBI_TaxID=111955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN JCM 10545 / 7;
RX   PubMed=11572479;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA   Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA   Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA   Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA   Oshima T., Kikuchi H.;
RT   "Complete genome sequence of an aerobic thermoacidophilic
RT   Crenarchaeon, Sulfolobus tokodaii strain7.";
RL   DNA Res. 8:123-140(2001).
DR   EMBL; AP000984; BAB66041.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   531 AA;  58083 MW;  25B3DC3D00841658 CRC64;


  Query Match             80.0%;  Score 36;  DB 17;  Length 531;
  Best Local Similarity   66.7%;  Pred. No. 26;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
           |:| ||:||
Db     102 NIFSTPILA 110


RESULT   4
Q9LZ27
ID   Q9LZ27      PRELIMINARY;      PRT;   669 AA.
AC   Q9LZ27;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 75.3 KDA PROTEIN (GENOMIC DNA, CHROMOSOME 5, P1
DE   CLONE:MUK11).
GN   T1E3_60.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II, Brassicales, Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Terryn N., Ardiles W., Buysshaert C., Dasseville R.,
RA   De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA   Villaroel R., Gielen J., Van Montagu M., Bancroft I., Mewes H.W.,
RA   Rudd S., Lemcke K., Mayer K.F.X.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=98162728; PubMed=9501997;
RA   Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT   Sequence features of the regions of 1,191,918 bp covered by seventeen
RT   physically assigned P1 clones.";
RL   DNA Res. 4:401-414(1997).
DR   EMBL; AL162972; CAB86012.1; -.
DR   EMBL; AB008271; BAB08974.1; -.
DR   InterPro; IPR002110; ANK.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   ANK repeat; Hypothetical protein; Repeat.
SQ   SEQUENCE   669 AA;  75339 MW;  676ACC5795885184 CRC64;


  Query Match             80.0%;  Score 36;  DB 10;  Length 669;
  Best Local Similarity   75.0%;  Pred. No. 32;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFEIPIL 8
           ||:|||:|
Db     112 NLYETPLL 119


RESULT   5
Q9LZ29
ID   Q9LZ29      PRELIMINARY;      PRT;   705 AA.
AC   Q9LZ29;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01 OCT 2000 (TrEMBLrel. 15, Last sequence update)
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 79.5 KDA PROTEIN.
GN   T1E3_40.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC   Spermatophyta, Magnoliophyta, Eudicotyledons, Core eudicots, Rosidae,
OC   eurosids II; Brassicales, Brassicaceae, Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Terryn N., Ardiles W., Buysshaert C., Dasseville R.,
RA   De Clercq R., De Keyser A., Neyt P., Poure F., Van Den Daele H.,
RA   Villarroel R., Giglen J., Van Montagu M., Pancroft I., Mewes H.W.,
RA   Rudd S., Lemcke K., Mayer K.F.X.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL162972; CAB86010.1; -.
DR   InterPro; IPR002110; ANK.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   ANK repeat; Hypothetical protein; Repeat.
SQ   SEQUENCE   705 AA;  79491 MW;  91C44E144E01ZCC7 CRC64;


  Query Match             80.0%;  Score 56;  DB 18;  Length 705;
  Best Local Similarity   75.0%;  Pred. No. 34;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPIL 8
           ||:|||:|
Db      99 NLYETPLL 106


RESULT   6
O06544
ID   O06544      PRELIMINARY;      PRT;   250 AA.
AC   O06544;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PROBABLE 3 HYDROXYACYL-COA DEHYDROGENASE 25.8 KDA TYPE II
DE   (EC 1.1.1.35) (TYPE II HADH)
GN   RV1144 OR MTCI65.11
OS   Mycobacterium tuberculosis.
OC   Bacteria, Firmicutes, Actinobacteria, Actinobacteridae,
OC   Actinomycetales, Corynebacterineae, Mycobacteriaceae, Mycobacterium
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- CATALYTIC ACTIVITY: L-3-HYDROXYACYL-CoA + NAD(+) = 3-OXOACYL-COA +
CC       NADH.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       FAMILY (SDR).
DR   EMBL; Z95584; CAB09032.1; -.
DR   HSSP; O70351; 1E3S.
DR   TubercuList; Rv1144; -.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00106; adh_short; 1.
DR   PROSITE; PS00061; ADH_SHORT; 1.
KW   Hypothetical protein; Oxidoreductase; NAD; Mitochondrion;
KW   Complete proteome.
FT   DOMAIN       32     35       POLY-VAL.
FT   NP_BIND       7     32       NAD (BY SIMILARITY).
FT   ACT_SITE    156    156       BY SIMILARITY.
SQ   SEQUENCE   250 AA;  25787 MW;  123A1805A1/B0BE CRC64;


  Query Match             75.0%;  Score 34;  DB 16;  Length 250;
  Best Local Similarity   75.0%;  Pred. No. 32;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LFETPILA 9
           |:|||:||
Db     188 LFDTPLLA 195


RESULT   7
Q9RAH3
ID   Q9RAH3      PRELIMINARY;      PRT;  1244 AA.
AC   Q9RAH3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NOSB.
GN   NOSB.
OS   Nostoc sp. GSV224.
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=76334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GSV224;
RA   Hoffmann D., Hevel J.M., Moore R.E.;
RT   "Characterization of the nostopeptolide biosynthetic gene cluster of
RT   Nostoc sp. GSV224.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF204805; AAF15892.2; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227, Acyltransf_domain
DR   InterPro; IPR000794, Beta-ketoacyl-synth.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR003880, Phosphopant_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; Ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_1.
DR   PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE  1244 AA;  137549 MW;  A9BCF74288A8A881 CRC64;


  Query Match             77.6%;  Score 34;  DB 2;  Length 1244;
  Best Local Similarity   66.7%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
           ||||:| :|
Db    1210 NLFESPTIA 1218


RESULT   8
Q75814
ID   Q75814      PRELIMINARY;      PRT;   245 AA.
AC   Q75814;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
```

```
DE   REVERSE TRANSCRIPTASE (FRAGMENT).
GN   POL.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TR324 FROM AUSTRALIA;
RA   Zhong N.N., Harren L., Nellan B.A., Cooper D.A., Delaney S.F.,
RA   McQueen P.W.;
RT   "Sequence analyses of the reverse transcriptase region of HIV-1
RT   isolates from Sydney, Australia.";
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U64171; AAB05314.1; -.
DR   InterPro; IPR000477; RVTse
DR   Pfam; PF00078; rvt; 1.
KW   RNA-directed DNA polymerase.
FT   NON_TER       1       1
FT   NON_TER     245     245
SQ   SEQUENCE   245 AA;  28756 MW;  6D51EA33BFC3E486 CRC64;


  Query Match             73.3%;  Score 33;  DB 15;  Length 245;
  Best Local Similarity   66.7%;  Pred. No. 51;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NLFETPILA 9
           ||: ||| |
Db      44 NLYNTPIFA 52


RESULT   9
Q92AH5
ID   Q92AH5      PRELIMINARY;      PRT;   304 AA.
AC   Q92AH5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PYRD PROTEIN.
GN   PYRD OR LIN1947.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / SEROVAR 6A;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596170; CAC97177.1; -.
DR   ListiList; LIN01947; -.
KW   Complete proteome.
SQ   SEQUENCE   304 AA;  32225 MW;  67391413A4C85291 CRC64;


  Query Match             73.3%;  Score 33;  DB 16;  Length 304;
  Best Local Similarity   85.7%;  Pred. No. 64;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 FETPILA 9
           |||||:|
Db      92 FETPIIA 98


RESULT  10
O44508
ID   O44508      PRELIMINARY;      PRT;   572 AA.
AC   O44508;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 66.6 KDA PROTEIN.
GN   F42G8.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=98068913; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans, a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Gattung S., Holmes A.;
RT   "The sequence of C. elegans cosmid F42G8.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF038618, AAB92073.1; -.
DR   InterPro; IPR001763; Rhodanese_domain.
DR   InterPro; IPR000063; Thiored.
DR   SMART; SM00450; RHOD; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   572 AA;  66648 MW;  5B1C8EF2343B474C CRC64;


  Query Match             73.3%;  Score 33;  DB 5;  Length 572;
  Best Local Similarity   75.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LFETPILA 9
           ::|||||
Db     455 VYETPILA 462


RESULT  11
O74798
ID   O74798      PRELIMINARY;      PRT;   658 AA.
AC   O74798;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL 57.9 KDA PROTEIN C2D10.04 IN CHROMOSOME II.
GN   SPBC2D10.04.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Wood V., Rajandream M.A., Barrell B.G., Taylor K., Harris D.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO YEAST YJL084L.
```

```
DR   EMBL; AL031788; CAA21162.1; -.
SQ   SEQUENCE   658 AA;  72731 MW;  49FB066BB970B4F7 CRC64;


  Query Match             73.3%;  Score 33;  DB 3;  Length 658;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NLFETP 6
          ||||||
Db    423 NLFETP 428


RESULT  12
Q21225
ID   Q21225      PRELIMINARY;      PRT;   694 AA.
AC   Q21225;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   K04G2.6 PROTEIN.
GN   K04G2.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Gardner A.E.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z75712; CAB00043.1; -.
SQ   SEQUENCE   694 AA;  78366 MW;  9F4655D087A3D49B CRC64;


  Query Match             73.3%;  Score 33;  DB 5;  Length 694;
  Best Local Similarity   66.7%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 NLFETPILA 9
          || |:|:||
Db    584 NLVESPVLA 592


RESULT  13
Q9BV54
ID   Q9BV54      PRELIMINARY;      PRT;  1162 AA.
AC   Q9BV54;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 128.7 KDA PROTEIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA, CHORIOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (DEC 2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC001474; AAH01474.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Helicase; Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   1162 AA;  128656 MW;  42FE1C6144E7BBB6 CRC64;


  Query Match             73.3%;  Score 33;  DB 4;  Length 1162;
  Best Local Similarity   75.0%;  Pred. No. 2.5e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 NLFETPIL 8
          |:|| |||
Db    217 NMFERPIL 224


RESULT  14
Q9Y4B4
ID   Q9Y4B4      PRELIMINARY;      PRT;  1241 AA.
AC   Q9Y4B4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA0809 PROTEIN (FRAGMENT).
GN   KIAA0809.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99087487; PubMed=9872452;
RA   Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XI.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 5:277-286(1998).
DR   EMBL; AB018352; BAA34529.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Helicase.
FT   NON_TER       1       1
SQ   SEQUENCE   1241 AA;  137451 MW;  E7B11A45F8769C40 CRC64;


  Query Match             73.3%;  Score 33;  DB 4;  Length 1241;
  Best Local Similarity   75.0%;  Pred. No. 2.7e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 NLFETPIL 8
          |:|| |||
Db    296 NMFERPIL 303


RESULT  15
Q99NG0
ID   Q99NG0      PRELIMINARY;      PRT;  1466 AA.
AC   Q99NG0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   STEROID RECEPTOR-INTERACTING SNF2 DOMAIN PROTEIN.
GN   SRISNF2L OR SRISNF2L.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN SWISS WEBSTER /NIH; TISSUE=11-DAY EMBRYO;
RA   Rouleau N., Reeben M., Moilanen A.M., Palvimo J.J., Janne O.A.;
RT   "Identification of a novel SNF2-like family member protein as a
RT   coregulator in steroid receptor-mediated gene transcription.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ132389; CAC24704.1; -.
DR   MGD; MGI:1933196; Srisnf21.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Helicase; Receptor.
SQ   SEQUENCE   1466 AA;  162539 MW;  D5945AD802B93D12 CRC64;
```

Query Match 73.4%; Score 33; DB 11; Length 1466;
Best Local Similarity 75.0%; Pred. No. 3.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 NLFEIPIL 8
           |:|| |||
Db     521 NMFERPIL 528
```

Search completed: September 5, 2002, 15:31:03
Job time: 468 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:27:34 ; Search time 58.86 Seconds
(without alignments)
3.735 Million cell updates/sec

Title: US-09-744-804-38
Perfect score: 46
Sequence: 1 NLFETPVEA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	46	100.0	160	2	US-08-162-402B-10	Sequence 10, Appl
2	46	100.0	217	1	US-07-607-538C-3	Sequence 3, Appli
3	46	100.0	217	2	US-08-162-402B-3	Sequence 3, Appli
4	46	100.0	218	1	US-07-607-538C-2	Sequence 2, Appli
5	46	100.0	218	2	US-08-162-402B-2	Sequence 2, Appli
6	46	100.0	387	2	US-08-162-402B-6	Sequence 6, Appli
7	46	100.0	465	2	US-08-162-402B-8	Sequence 8, Appli
8	37	80.4	159	2	US-08-162-402B-12	Sequence 12, Appl
9	36	78.3	14	2	US-08-162-402B-26	Sequence 26, Appl
10	32	69.6	406	4	US-08-584-008A-8	Sequence 8, Appli
11	31	67.4	108	4	US-09-187-859-40	Sequence 40, Appl
12	31	67.4	162	4	US-08-936-165A-344	Sequence 344, App
13	31	67.4	217	3	US-08-971-158-1	Sequence 1, Appli
14	31	67.4	217	3	US-08-971-158-3	Sequence 3, Appli
15	30	65.2	127	4	US-09-437-054A-12	Sequence 12, Appl
16	30	65.2	391	4	US-08-984-618-11	Sequence 11, Appl
17	30	65.2	799	4	US-09-136-390-6	Sequence 6, Appli
18	30	65.2	857	2	US-08-659-251-2	Sequence 2, Appli
19	30	65.2	857	4	US-09-256-448-2	Sequence 2, Appli
20	30	65.2	857	5	PCT-US96-11445-2	Sequence 2, Appli
21	30	65.2	1078	4	US-08-922-435-22	Sequence 22, Appl
22	29	63.0	58	2	US-08-700-152A-2	Sequence 2, Appli
23	29	63.0	68	3	US-08-912-272-5	Sequence 5, Appli
24	29	63.0	68	4	US-09-026-039-5	Sequence 5, Appli
25	29	63.0	79	3	US-08-904-263A-6	Sequence 6, Appli
26	29	63.0	181	4	US-08-858-207A-317	Sequence 317, App
27	29	63.0	188	2	US-08-385-590A-6	Sequence 6, Appli
28	29	63.0	188	2	US-08-385-590A-28	Sequence 28, Appl
29	29	63.0	188	2	US-08-921-520-6	Sequence 6, Appli
30	29	63.0	188	4	US-09-021-520-28	Sequence 28, Appl
31	29	63.0	330	2	US-08-487-031-10	Sequence 10, Appl
32	29	63.0	330	3	US-08-473-034-10	Sequence 10, Appl
33	29	63.0	442	2	US-08-700-152A-4	Sequence 4, Appli
34	29	63.0	802	4	US-09-147-236-4	Sequence 4, Appli
35	29	63.0	1231	3	US-08-904-263A-4	Sequence 4, Appli
36	29	63.0	1805	1	US-07-854-913-2	Sequence 2, Appli
37	29	63.0	3077	6	5223423 2	Patent No. 5223423
38	29	63.0	3135	1	US-08-323-170B-2	Sequence 2, Appli
39	29	63.0	3135	4	US-08-954-441-2	Sequence 2, Appli
40	29	63.0	3418	2	US-08-639-501-2	Sequence 2, Appli
41	29	63.0	3418	2	US-08-603-753D-4	Sequence 4, Appli
42	29	63.0	3418	4	US-09-044-946-2	Sequence 2, Appli
43	29	63.0	3418	3	US 08 755 587 44	Sequence 44, Appl
44	29	63.0	3418	3	US-09-044-908-2	Sequence 2, Appli
45	29	63.0	3418	4	US 09 099 753 4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-162-402B-10
; Sequence 10, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS.
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA.
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Assel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION.
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 160 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-10

Query Match 100.0%, Score 46, DB 2, Length 160;
Best Local Similarity 100.0%, Pred. No. 0.038;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NLFETPVEA 9
           |||||||||
Db     126 NLFETPVEA 134
```

```
RESULT   2
US-07-607-538C-3
; Sequence 3, Application US/07607538C
; Patent No. 5455031
;  GENERAL INFORMATION:
;    APPLICANT:  Ceriani Dr., Roberto L.
;    APPLICANT:  Peterson Dr., Jerry A.
;    APPLICANT:  Larocca, David J.
;    TITLE OF INVENTION:  POLYPEPTIDE WITH 46
;    TITLE OF INVENTION:  DIFFERENTIATION ANTIGEN BINDING SPECIFITY AND  CLOTTING
;    TITLE OF INVENTION:  FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
;    TITLE OF INVENTION:  FUSION  PROTEIN,  POLYNUCLEOTIDE AND  POLYRIBO-
;    TITLE OF INVENTION:  NUCLEOTIDE  ENCODING  THE POLYPEPTIDE, ANTI-
;    TITLE OF INVENTION:  POLYPEPTIDE ANTIBODIES, KITS  AND  METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  V.  Amzel & Assoc
;      STREET:  2055 No. 5455031th Broadway
;      CITY:  Walnut Creek
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94596
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS 5.0
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/607,538C
;      FILING DATE:  01-NOV-1990
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Viviana Amzel
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  CRFCC-004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 943-1931
;      TELEFAX:  (510) 943-1189
;      TELEX:  N.A.
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  217 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:
US-07-607-538C-3
```

Query Match 100.0%; Score 46; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.053;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NLFETPVEA 9
           |||||||||
Db      24 NLFETPVEA 32
```

```
RESULT   3
US-08-162-402B-3
; Sequence 3, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  217 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-3
```

Query Match 100.0%; Score 46; DB 2; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.053;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NLFETPVEA 9
           |||||||||
Db      24 NLFETPVEA 32
```

```
RESULT   4
US-07-607-538C-2
; Sequence 2, Application US/07607538C
; Patent No. 5455031
;  GENERAL INFORMATION:
;    APPLICANT:  Ceriani Dr., Roberto L.
;    APPLICANT:  Peterson Dr., Jerry A.
;    APPLICANT:  Larocca, David J.
;    TITLE OF INVENTION:  POLYPEPTIDE WITH 46
;    TITLE OF INVENTION:  DIFFERENTIATION ANTIGEN BINDING SPECIFITY AND  CLOTTING
;    TITLE OF INVENTION:  FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
;    TITLE OF INVENTION:  FUSION  PROTEIN,  POLYNUCLEOTIDE AND  POLYRIBO-
;    TITLE OF INVENTION:  NUCLEOTIDE  ENCODING  THE POLYPEPTIDE, ANTI-
;    TITLE OF INVENTION:  POLYPEPTIDE ANTIBODIES, KITS  AND  METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  V.  Amzel & Assoc.
;      STREET:  2055 No. 5455031th Broadway
;      CITY:  Walnut Creek
;      STATE:  California
;      COUNTRY:  USA
```

```
;       ZIP:  94596
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS 5.0
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/607,538C
;       FILING DATE:  01-NOV-1990
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Viviana Amzel
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  CRB00-004
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (510) 943-1931
;       TELEFAX:  (510) 943-1189
;       TELEX:  N.A.
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  218 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     FRAGMENT TYPE:
US-07-607-538C-2


  Query Match             100.0%;  Score 46;  DB 1;  Length 218;
  Best Local Similarity   100.0%;  Pred. No. 0.053;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |||||||||
Db      25 NLFETPVEA 33


RESULT   5
US-08-162-402B-2
; Sequence 2, Application US/08/162,402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;     APPLICANT:  CERIANI, ROBERTO L.
;     APPLICANT:  PETERSON, JERRY A.
;     APPLICANT:  LAROCCA, DAVID J.
;     TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;     TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;     NUMBER OF SEQUENCES:  29
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pretty, Schroeder & Poplawski
;       STREET:  444 South Flower St., 19th Floor
;       CITY:  Los Angeles
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/162,402B
;       FILING DATE:  03-DEC-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Amzel, Viviana
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  P66 38215
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  213-622-7700
;       TELEFAX:  213 489-4210
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  218 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
US-08-162-402B-2


  Query Match             100.0%;  Score 46;  DB 2;  Length 218;
  Best Local Similarity   100.0%;  Pred. No. 0.053;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |||||||||
Db      25 NLFETPVEA 33


RESULT   6
US-08-162-402B-6
; Sequence 6, Application US/08/162,402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;     APPLICANT:  CERIANI, ROBERTO L.
;     APPLICANT:  PETERSON, JERRY A.
;     APPLICANT:  LAROCCA, DAVID J.
;     TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;     TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;     NUMBER OF SEQUENCES:  29
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pretty, Schroeder & Poplawski
;       STREET:  444 South Flower St., 19th Floor
;       CITY:  Los Angeles
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/162,402B
;       FILING DATE:  03-DEC-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Amzel, Viviana
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  P66 38215
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  213 622 7700
;       TELEFAX:  213-489-4210
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  387 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
US-08-162-402B-6


  Query Match             100.0%;  Score 46;  DB 2;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 NLFETPVEA 9
           |||||||||
Db     194 NLFETPVEA 202


RESULT   7
US-08-162-402B-8
; Sequence 8, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St , 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  465 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B 8


  Query Match             100.0%;  Score 46;  DB 2;  Length 465;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |||||||||
Db     272 NLFETPVEA 280


RESULT   8
US-08-162-402B-12
; Sequence 12, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  159 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-12


  Query Match             80.4%;  Score 37;  DB 2;  Length 159;
  Best Local Similarity   77.8%;  Pred. No. 2.4;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |||||| :|
Db     128 NLFETPILA 136


RESULT   9
US-08-162-402B-26
; Sequence 26, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
```

```
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B 26


  Query Match             78.3%;  Score 36;  DB 2;  Length 14;
  Best Local Similarity   75.0%;  Pred. No. 0.24;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
            |||:||:|
Db       2 NLFDTPLE 9


RESULT  10
US-08-584-008A-8
; Sequence 8, Application US/08584008A
; Patent No. 6277615
;  GENERAL INFORMATION:
;    APPLICANT:  Varghese, Joseph N.
;    APPLICANT:  Garrett, Thomas P.J.
;    APPLICANT:  Fincher, Geoffrey B.
;    APPLICANT:  Hoj, Peter B.
;    APPLICANT:  Chen, Lin
;    TITLE OF INVENTION:  BETA-GLUCANASE OF ENHANCED STABILITY
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JACOBSON, PRICE, HOLMAN & STERN
;      STREET:  The Jenifer Building, 400 Seventh St. N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/584,008A
;      FILING DATE:  11-JAN-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PL9821
;      FILING DATE:  07-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU94/00377
;      FILING DATE:  06-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Holman, J.C.
;      REGISTRATION NUMBER:  22,769
;      REFERENCE/DOCKET NUMBER:  9944/P60173
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-638-6666
;      TELEFAX:  202-393-5350
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  306 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-584-008A-8


  Query Match             69.6%;  Score 32;  DB 4;  Length 306;
  Best Local Similarity   66.7%;  Pred. No. 51;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
            |||:| |:|
Db     205 NLFDTIVDA 213


RESULT  11
US 09 187 859-40
; Sequence 40, Application US/0918/859A
, Patent No. 6358920
, GENEFAL INFORMATION:
;  APPLICANT: Blaschuk, Orest W.
,  APPLICANT: Gour, Barbara J.
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
;  TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
;  FILE REFERENCE: 100086.407C1
;  CURRENT APPLICATION NUMBER: US/09/187,859A
;  CURRENT FILING DATE: 1998-11-06
;  NUMBER OF SEQ ID NOS: 4052
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40
;   LENGTH: 108
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-187-859-40


  Query Match             67.4%;  Score 31;  DB 4;  Length 108;
  Best Local Similarity   75.0%;  Pred. No. 25;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LFETPVEA 9
            |:| ||||
Db      73 LYEIPVEA 80


RESULT  12
US-08-936-165A-344
; Sequence 344, Application US/08936165A
; Patent No. 6348582
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael
;    APPLICANT:  Burnham, Martin
;    APPLICANT:  Hodgson, John
;    APPLICANT:  Knowles, David
,    APPLICANT:  Lonetto, Michael
;    APPLICANT:  Nicholas, Richard
;    APPLICANT:  Pratt, Julie
;    APPLICANT:  Reichard, Richard
;    APPLICANT:  Rosenberg, Martin
;    APPLICANT:  Ward, Judith
,    TITLE OF INVENTION:  No. 6348582A1 Prokaryotic Polypeptides,
,    TITLE OF INVENTION:  Polypeptides and Their Uses
;    NUMBER OF SEQUENCES:  534
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
,      COUNTRY:  USA
;      ZIP:  19406-0939
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/936,165A
;      FILING DATE:  24-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/027,032
;      FILING DATE:  24-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  PA0549
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  344:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  162 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Protein
US-08-936-165A-344


  Query Match             67.4%;  Score 31;  DB 4;  Length 162;
  Best Local Similarity   62.5%;  Pred No. 39;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           | |:|| |
Db     117 NFFDTPAE 124


RESULT  13
US-08-971-158-1
; Sequence 1, Application US/08971158
; Patent No. 6010879
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Goli, Surya K
;    TITLE OF INVENTION:  HUMAN MITOCHONDRIAL CHAPERONE PROTEIN
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/971,158
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/824,875
;      FILING DATE:  March 26, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0245-1 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  217 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  COLNNOT08
;      CLONE:  1844121
US-08-971-158-1


  Query Match             67.4%;  Score 31;  DB 3;  Length 217;
  Best Local Similarity   85.7%;  Pred No. 55;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LFETPVE 8
             || ||||
Db     178 LFHTPVE 184


RESULT  14
US-08-971-158-3
; Sequence 3, Application US/08971158
; Patent No. 6010879
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  HUMAN MITOCHONDRIAL CHAPERONE PROTEIN
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/971,158
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/824,875
;      FILING DATE:  March 26, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0245-1 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  217 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  1695878
US-08-971-158-3


  Query Match             67.4%;  Score 31;  DB 3;  Length 217;
  Best Local Similarity   85.7%;  Pred No. 55;
```

Matches 6; Conservative 0, Mismatches 1, Indels 0, Gaps 0,

```
Qy       2 LFETPVE 8
           || |||| 
Db     178 LFHTPVE 184
```

RESULT 15
US-09-437-054A-12
; Sequence 12, Application US/09437054A
; Patent No. 6316698
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Kinney, Anthony J.
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
; FILE REFERENCE: BB1273 US NA
; CURRENT APPLICATION NUMBER: US/09/437,054A
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/107,909
; PRIOR FILING DATE: 1998-No. 6316698ember-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Zea mays
US-09-437-054A-12

Query Match 65.2%; Score 30, DB 4, Length 127,
Best Local Similarity 62.5%; Pred. No. 47,
Matches 5; Conservative 1, Mismatches 2, Indels 0, Gaps 0,

```
Qy       1 NLFETPVE 8
           | | |||: 
Db      55 NFFRTPVD 62
```

Search completed: September 5, 2002, 15:27:35
Job time: 320 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:26:38 ; Search time 154.44 Seconds
(without alignments)
6.310 Million cell updates/sec

Title: US-09-744-804-38
Perfect score: 46
Sequence: 1 NIFETPVEA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	46	100.0	9	21	AAY82843	Lactadherin (BA-46
2	46	100.0	217	16	AAR77253	HMFG 46 kDa antige
3	46	100.0	218	16	AAR77254	HMFG 46 kDa antige
4	46	100.0	387	16	AAR77252	HMFG 46 kDa antige
5	46	100.0	387	21	AAY94453	Human lactadherin
6	38	82.6	26	22	AAR62694	ABC1 protein exter
7	38	82.6	1089	22	AAR62691	Human ABC1 protein
8	38	82.6	1144	22	AAU02184	Human ABC1 mutant
9	38	82.6	1144	22	AAU02185	Human ABC1 mutant
10	38	82.6	1525	22	AAU02187	Human ABC1 mutant
11	38	82.6	2130	22	AAU02190	Human ABC1 mutant
12	38	82.6	2143	21	AAB38108	Human ABC1 cholest
13	38	82.6	2201	21	AAY79380	Human ATP binding
14	38	82.6	2201	22	AAE13021	Human ATP binding
15	38	82.6	2201	22	AAM50227	Human ATP binding
16	38	82.6	2259	21	AAB38107	Human ABC1 FHA-3 m
17	38	82.6	2260	21	AAB38106	Human ABC1 cholest
18	38	82.6	2261	21	AAB38082	Human ABC1 cholest
19	38	82.6	2261	21	AAB38104	Human ABC1 cholest
20	38	82.6	2261	21	AAB38105	Human ABC1 cholest
21	38	82.6	2261	21	AAB38109	Human ABC1 cholest
22	38	82.6	2261	21	AAB38110	Human ABC1 cholest
23	38	82.6	2261	21	AAB38111	Human ABC1 cholest
24	38	82.6	2261	21	AAB38112	Human ABC1 cholest
25	38	82.6	2261	21	AAB38113	Human ABC1 cholest
26	38	82.6	2261	21	AAB38114	Human ABC1 cholest
27	38	82.6	2261	21	AAB38115	Human ABC1 cholest
28	38	82.6	2261	21	AAB38116	Human ABC1 cholest
29	38	82.6	2261	21	AAB38117	Human ABC1 cholest
30	38	82.6	2261	22	AAE13022	Human ATP binding
31	38	82.6	2261	22	AAM50228	Human ATP binding
32	38	82.6	2261	22	AAM78510	Human protein SEQ
33	38	82.6	2261	22	AAU02176	Human ABC1 Homo
34	38	82.6	2261	22	AAU02177	Human ABC1 mutant
35	38	82.6	2261	22	AAU02181	Human ABC1 mutant
36	38	82.6	2261	22	AAU02182	Human ABC1 mutant
37	38	82.6	2261	22	AAU02183	Human ABC1 mutant
38	38	82.6	2261	22	AAU02186	Human ABC1 mutant
39	38	82.6	2261	22	AAU02188	Human ABC1 mutant
40	38	82.6	2261	22	AAU02189	Human ABC1 mutant
41	38	82.6	2261	22	AAB71749	Human ABC1 protein
42	38	82.6	2261	22	AAB31361	Amino acid sequenc
43	38	82.6	2261	22	AAB31362	Amino acid sequenc
44	38	82.6	2261	22	AAB31363	Amino acid sequenc
45	38	82.6	2261	22	AAB31365	Amino acid sequenc

ALIGNMENTS

```
RESULT 1
AAY82843
ID   AAY82843 standard; peptide; 9 AA.
XX
AC   AAY82843;
XX
DT   19-JUN-2000 (first entry)
XX
DE   Lactadherin (BA-46) peptide fragment (tumour associated antigen).
XX
KW   Tumour associated antigen peptide; TAA; cancer; carcinoma;
KW   treatment; prevention; cure; anti-tumour vaccine; metastases;
KW   breast; bladder; prostate; pancreas; ovary; thyroid; colon;
KW   stomach; carcinoma; MHC Class I; HLA A2; human;
KW   Major Histocompatibility Complex; uroplakin;
KW   prostate specific antigen; prostate specific membrane antigen;
KW   prostate acid phosphatase; mucin; lactadherin;
KW   teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRIPTO-1.
XX
OS   Homo sapiens.
XX
PN   WO200006723-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-IL00417.
XX
PR   30-JUL-1998;   98IL-0125608.
XX
PA   (YEDA ) YEDA RES & DEV CO LTD.
PA   (BIOT-) BIO-TECHNOLOGY GEN CORP.
XX
PI   Eisenbach L, Carmon L, Tirosh B, Bar haim E, Paz A, Fridkin M;
PI   Fitzer-attas C;
```

XX
DR WPI; 2000-205463/18.
XX
PT Tumor associated antigen peptides, especially derived from uroplakin, useful as vaccines to prevent or cure cancers including breast, bladder, prostate, pancreas, ovary, thyroid, colon and stomach -
XX
PS Claim 17; Page 100; 114pp; English.
XX
CC Tumour associated antigen peptides (TAA) may be used for the treatment, prevention and cure of cancer or cancer metastases. The cancer may be breast, bladder, prostate, pancreas, ovary, thyroid, colon, stomach, head or neck cancer or a carcinoma. The tumour associated antigens are presentable to the immune system by HLA-A2 molecules and are generally between 8 to 10 amino acids in length. The amino acids located at positions 2 and 9 of the tumour associated antigens are the anchor residues which participate in the binding to MHC class I molecules, more specifically HLA-A2. More tumour associated antigens are described in CENESEQ records AAY82806-Y82882. Those tumour associated antigens described in records AAY82806-Y82824 and AAY82855-YY82869 are derived from Uroplakin, such as Uroplakin II, Uroplakin Ia, Uroplakin III and Uroplakin Ib. Those described in records AAY82825-Y82829 are derived from prostate specific antigen (PSA). Those described in records AAY82830-Y82835 are derived from prostate specific membrane antigen (PSMA). Those described in records YY82836 AAY82839 are derived from prostate acid phosphatase (PAP). Those described in records AAY82840-Y82846 are derived from Lactadherin (BA-46). Those described in records AAY82847-Y82854 are derived from Mucin and those described in records AAY82871-Y82882 are derived from Teratocarcinoma derived growth factor (CRIPTO-1).
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 46; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPVEA 9
||||||||| |
Db 1 nlfetpvea 9

RESULT 2
AAR77253
ID AAR77253 standard; Protein; 217 AA.
XX
AC AAR77253;
XX
DT 21-NOV-1995 (first entry)
XX
DE HMFG 46 kDa antigen C-terminal region.
XX
KW HMFG, human milk fat globule; antigen; immunogen; vaccine; virucide; KW epithelium; tumor; breast cancer; monoclonal antibody; MAb.
XX
OS Homo sapiens.
XX
PN WO9515171-A.
XX
PD 08-JUN-1995.
XX
PF 05-DEC-1994; 94WO-US13967.
XX
PR 03-DEC-1993; 93US-0162402.
XX
PA (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI Ceriani RL, Larocca DJ, Peterson JA;
XX
DR WPI; 1995-215151/28.
XX
PT 46 kD apparent molecular weight human milk fat globule antigen -
PT used in assays to determine the presence of a cancerous tumour of
PT epithelial origin, and in a vaccine against neoplastic tumours
XX
PS Claim 7; Page 41; 68pp; English.
XX
CC A partial cDNA clone BA46-1 (AAQ91199) for the 46 kDa HMFG antigen, a major component of the apical surface of the normal breast epithelial cell, was obtained by PCR of cDNA from a lactating breast cDNA library. The C-terminal region of the encoded protein showed 43% identity to corresp. regions of human Factor-V and 38% to Factor-VIII
XX
SQ Sequence 217 AA;

Query Match 100.0%; Score 46; DB 16; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPVEA 9
||||||||| |
Db 24 nlfetpvea 32

RESULT 3
AAR77254
ID AAR77254 standard; Protein; 218 AA.
XX
AC AAR77254;
XX
DT 21-NOV-1995 (first entry)
XX
DE HMFG 46 kDa antigen partial sequence.
XX
KW HMFG, human milk fat globule; antigen; immunogen; vaccine; virucide; KW epithelium; tumor; breast cancer; monoclonal antibody; MAb.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 69..71
FT /label= N-glycosylation_site
FT Modified-site 156..158
FT /label= N-glycosylation_site
FT Modified-site 160..162
FT /label= N-glycosylation_site
FT Modified-site 181..183
FT /label= N-glycosylation_site
XX
PN WO9515171-A.
XX
PD 08-JUN-1995.
XX
PF 05-DEC-1994; 94WO-US13967.
XX
PR 03-DEC-1993; 93US-0162402.
XX
PA (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI Ceriani RL, Larocca DJ, Peterson JA;
XX
DR WPI; 1995-215151/28.
DR N-PSDB; AAQ91199.
XX
PT 46 kD apparent molecular weight human milk fat globule antigen -
PT used in assays to determine the presence of a cancerous tumour of
PT epithelial origin, and in a vaccine against neoplastic tumours
XX
PS Example 7; Page 38-40; 68pp; English.
XX
CC A partial cDNA clone BA46-1 (AAQ91199) for the 46 kDa HMFG antigen, a major component of the apical surface of the normal breast epithelial cell, was obtained by PCR of cDNA from a lactating

```
CC   breast cDNA library.
XX
SQ   Sequence   218 AA;


  Query Match             100.0%;  Score 46;  DB 16;  Length 218;
  Best Local Similarity   100.0%;  Pred. No. 0.19;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |||||||||
Db      25 nlfetpvea 33


RESULT   4
AAR77252
ID   AAR77252 standard; Protein; 387 AA.
XX
AC   AAR77252;
XX
DT   21-NOV-1995  (first entry)
XX
DE   HMFG 46 kDa antigen.
XX
KW   HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
KW   epithelium; tumor; breast cancer; monoclonal antibody; MAb.
XX
OS   Homo sapiens.
XX
PN   WO9515171-A.
XX
PD   08-JUN-1995.
XX
PF   05-DEC-1994;   94WO-US13967.
XX
PR   03-DEC-1993;   93US-0162402.
XX
PA   (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI   Ceriani RL,  Larocca DJ,  Peterson JA;
XX
DR   WPI; 1995-215151/28.
DR   N-PSDB; AAQ91198.
XX
PT   46 kD apparent molecular weight human milk fat globule antigen -
PT   used in assays to determine the presence of a cancerous tumour of
PT   epithelial origin, and in a vaccine against neoplastic tumours
XX
PS   Claim 6; Page 46-47, 68pp; English
XX
CC   A complete cDNA sequence for the 46 kDa HMFG antigen, a major
CC   component of the apical surface of the normal breast epithelial
CC   cell, was obtained by PCR and RACE methods.  cDNA clones can be
CC   used to prepare MAbs for use in immunotherapy, immunohistopathology,
CC   prognosis, diagnosis, imaging and therapy.  Recombinant antigen can
CC   be expressed in prokaryotic or (glycosylated) in eukaryotic cells.
XX
SQ   Sequence   387 AA;


  Query Match             100.0%;  Score 46;  DB 16;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.35;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |||||||||
Db     194 nlfetpvea 202


RESULT   5
AAY94453
ID   AAY94453 standard; Protein; 387 AA.
XX
AC   AAY94453;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Human lactadherin protein.
XX
KW   Human; lactadherin; MGF-E8; anti tumour; immune response;
KW   exosome; dendritic cell.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Secretion_signal
FT   Protein         24..387
FT                   /label= Lactadherin
FT   Binding site    46..48
FT                   /label= Integrin_binding_site
XX
PN   EP1004664-A1.
XX
PD   31-MAY-2000.
XX
PF   24-NOV-1998;   98EP-0402925.
XX
PR   24-NOV-1998;   98EP-0402925.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (CURI-) INST CURIE.
XX
DR   WPI; 2000-352597/31.
DR   N-PSDB; AAA27140.
XX
PT   Chimeric isolated (human) lactadherin polypeptide that functions as an
PT   adaptor of cross priming to eliminate pathogenic antigens, e.g. in
PT   tumors -
XX
PS   Example 3; Page 12; 20pp; English.
XX
CC   Lactadherin protein was found in exosomes produced by dendritic cells.
CC   The protein is involved in the phagocytosis of particulate antigens by
CC   dendritic cells. Exosomes produced by dendritic cells exposed to
CC   tumour antigens induce potent immune responses. Lactadherin or variants
CC   of it may be used in the mediation of an immune response. Variants of
CC   lactadherin may be used for inhibition and/or stimulation of the
CC   cross priming of antigens and stimulation of the phagocytosis of
CC   antigens by dendritic cells. Compositions derived from lactadherin can
CC   also be used to monitor an immune response, more specifically a CTL
CC   (cytotoxic T-lymphocyte) response and also to produce CTLs specific
CC   for a selected antigen. The present sequence is the human
CC   lactadherin protein.
XX
SQ   Sequence   387 AA;


  Query Match             100.0%;  Score 46;  DB 21;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.35;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |||||||||
Db     194 nlfetpvea 202


RESULT   6
AAB62694
ID   AAB62694 standard; peptide; 26 AA.
XX
AC   AAB62694;
XX
DT   06-AUG-2001  (first entry)
XX
DE   ABC1 protein external domain TM5-TM6 fragment (residues 795-820).
```

```
XX
KW   ABC1; antilipemic; cholesterol; inhibitor; low density lipoprotein; LDL.
XX
OS   Homo sapiens.
XX
PN   WO200132184-A2.
XX
PD   10-MAY-2001.
XX
PF   01-NOV-2000; 2000WO-US30109
XX
PR   01-NOV-1999;   99US-0162803.
PR   30-JUN-2000; 2000US-0215564.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Attie AD, Cook M, Gray-Keller MP, Hayden MR, Pimstone S;
PI   Brooks Wilson A,
XX
DR   WPI; 2001-335779/35
XX
PT   New method for inhibiting cholesterol uptake in the gut comprises
PT   administration of an inhibitor of an ABC1 protein -
XX
PS   Disclosure; Page 9, 41pp; English.
XX
CC   The invention relates to a new method for inhibiting cholesterol uptake
CC   in the gut that comprises administration of an inhibitor of an ABC1
CC   protein. The method is useful for: lowering levels of LDL (low density
CC   lipoprotein) cholesterol by reducing the activity of ABC1 protein in the
CC   intestinal cells and the abundance of the ABC1 protein in the individual
CC   by inhibiting the activity of the protein; identifying drugs that can
CC   lower serum cholesterol and LDL levels comprises assaying the drug to
CC   test if it can bind to an ABC1 protein; testing LDL cholesterol lowering
CC   agents; and for modulation of ABC1 biological activity. Sequences
CC   AAB62692-97 represent predicted external domain of ABC1 protein.
XX
SQ   Sequence   26 AA;


  Query Match             82.6%;  Score 38;  DB 22;  Length 26;
  Best Local Similarity   87.5%;  Pred. No. 0.82;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           ||||:|||
Db      14 nlfespve 21


RESULT   7
AAB62691
ID   AAB62691 standard; Protein; 1089 AA.
XX
AC   AAB62691;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Human ABC1 protein partial fragment
XX
KW   ABC1, antilipemic, cholesterol, inhibitor, low density lipoprotein; LDL.
XX
OS   Homo sapiens.
XX
PN   WO200132184-A2
XX
PD   10-MAY-2001
XX
PF   01-NOV-2000; 2000WO-US30109.
XX
PR   01-NOV-1999;   99US-0162803.
PR   30-JUN-2000; 2000US-0215564
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Attie AD, Cook M, Gray-Keller MP, Hayden MR, Pimstone S;
PI   Brooks-Wilson A;
XX
DR   WPI; 2001-335779/35.
DR   N-PSDB; AAF83826.
XX
PT   New method for inhibiting cholesterol uptake in the gut comprises
PT   administration of an inhibitor of an ABC1 protein -
XX
PS   Disclosure; Page 37-; 41pp; English.
XX
CC   The invention relates to a new method for inhibiting cholesterol uptake
CC   in the gut that comprises administration of an inhibitor of an ABC1
CC   protein. The method is useful for: lowering levels of LDL (low density
CC   lipoprotein) cholesterol by reducing the activity of ABC1 protein in the
CC   intestinal cells and the abundance of the ABC1 protein in the individual
CC   by inhibiting the activity of the protein, identifying drugs that can
CC   lower serum cholesterol and LDL levels comprises assaying the drug to
CC   test if it can bind to an ABC1 protein, testing LDL cholesterol lowering
CC   agents, and for modulation of ABC1 biological activity. The present
CC   sequence represents a human ABC1 protein partial fragment.
CC   Note: Pages containing part of the present sequence seems to be missing
CC   from the specification; the protein seems to contain at least 2261 amino
CC   acid residues as found by translating the corresponding DNA sequence.
XX
SQ   Sequence   1089 AA,


  Query Match             82.6%;  Score 38;  DB 22;  Length 1089;
  Best Local Similarity   87.5%;  Pred. No. 43;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           ||||:|||
Db     808 nlfespve 815


RESULT   8
AAU02184
ID   AAU02184 standard; Protein; 1144 AA.
XX
AC   AAU02184;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Human ABC1 mutant polypeptide #8.
XX
KW   Human; ABC1 gene; atherosclerosis, reverse transport, cholesterol;
KW   cardiovascular; neurological, Tangier disease,, LCAT deficiency; mutant;
KW   lecithin-cholesterol acetyltransferase; malaria, diabetes; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1114
FT                   /note= "Frameshift at amino acid 1114 introduces a stop
FT                         codon at amino acid 1144"
XX
PN   WO200130848-A2.
XX
PD   03-MAY-2001.
XX
PF   26-OCT-2000; 2000WO-EP10886.
XX
PP   26-OCT-1999;   99EP-0402668.
PR   01-MAR-2000; 2000US-0186260.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Denefle P, Kesler Mostas M, Arnould-Requigne I, Prades C, Naudin L;
PI   Lemoine C, Duverger N, Jaye M, Scarfoss CH, Remaley A, Brewer HB;
PI   Dean M;
```

```
XX
DR   WPI; 2001-316327/33.
DR   N-PSDB; AAS06132.
XX
PT   New human ABC1 nucleic acids and polypeptides for treating
PI   atherosclerosis, malaria and diabetes -
XX
PS   Claim 46; Page 281-287; 368pp; English.
XX
CC   The sequence represents the amino acid sequence of human ABC1 mutant
CC   polypeptide #8. The polypeptide and nucleic acid sequence, primers and
CC   probes derived from the ABC1 sequence, and vectors are useful for the
CC   prevention of atherosclerosis, in a subject affected by a dysfunction in
CC   the reverse transport of cholesterol. The polypeptide encoded by the ABC1
CC   gene is useful for screening for an active ingredient for the prevention
CC   or treatment of a disease resulting from dysfunction in the reverse
CC   transport of cholesterol. The nucleic acids and polypeptides are also
CC   useful for treating and preventing cardiovascular and neurological
CC   pathologies, and other diseases e.g. Tangier disease, lecithin-
CC   cholesterol (LCAT) deficiency, malaria and diabetes.
XX
SQ   Sequence   1144 AA;

  Query Match             82.6%;  Score 38;  DB 22;  Length 1144;
  Best Local Similarity   87.5%;  Pred. No. 45;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           ||||:|||
Db     808 nlfespve 815


RESULT   9
AAU02185
ID   AAU02185 standard; Protein; 1144 AA.
XX
AC   AAU02185;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Human ABC1 mutant polypeptide #9.
XX
KW   Human; ABC1 gene; atherosclerosis; reverse transport; cholesterol;
KW   cardiovascular; neurological; Tangier disease; LCAT deficiency; mutant;
KW   lecithin-cholesterol acetyltransferase; malaria; diabetes; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1114
FT                   /note= "Frameshift at amino acid 1114 introduces a stop
FT                          codon at amino acid 1144"
XX
PN   WO200130848-A2.
XX
PD   03-MAY-2001.
XX
PF   26-OCT-2000; 2000WO-EP10886.
XX
PR   26-OCT-1999;   99EP-0402668.
PR   01-MAR-2000; 2000US-0186260.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Denefle P,  Rosier-Montus M,  Arnould-Requique I,  Prades C,  Naudin L;
PI   Lemoine C,  Duverger N,  Jaye M,  Searfoss GH,  Romaley A,  Brewer HB;
PI   Dean M;
XX
DR   WPI; 2001-316327/33.
DR   N-PSDB; AAS06133.
XX
PT   New human ABC1 nucleic acids and polypeptides for treating
PT   atherosclerosis, malaria and diabetes -
XX
PS   Claim 46; Page 287 294; 368pp; English.
XX
CC   The sequence represents the amino acid sequence of human ABC1 mutant
CC   polypeptide #9. The polypeptide and nucleic acid sequence, primers and
CC   probes derived from the ABC1 sequence, and vectors are useful for the
CC   prevention of atherosclerosis, in a subject affected by a dysfunction in
CC   the reverse transport of cholesterol. The polypeptide encoded by the ABC1
CC   gene is useful for screening for an active ingredient for the prevention
CC   or treatment of a disease resulting from dysfunction in the reverse
CC   transport of cholesterol. The nucleic acids and polypeptides are also
CC   useful for treating and preventing cardiovascular and neurological
CC   pathologies, and other diseases e.g. Tangier disease, lecithin-
CC   cholesterol (LCAT) deficiency, malaria and diabetes.
XX
SQ   Sequence   1144 AA;

  Query Match             82.6%;  Score 38;  DB 22;  Length 1144;
  Best Local Similarity   87.5%;  Pred. No. 45;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           ||||:|||
Db     808 nlfespve 815


RESULT   10
AAU02187
ID   AAU02187 standard; Protein; 1525 AA.
XX
AC   AAU02187;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Human ABC1 mutant polypeptide #11.
XX
KW   Human; ABC1 gene; atherosclerosis; reverse transport; cholesterol;
KW   cardiovascular; neurological; Tangier disease; LCAT deficiency; mutant;
KW   lecithin-cholesterol acetyltransferase; malaria; diabetes; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1525
FT                   /note= "Wild type Trp substituted by STOP codon"
XX
PN   WO200130848-A2.
XX
PD   03-MAY-2001.
XX
PF   26-OCT-2000; 2000WO-EP10886.
XX
PR   26-OCT-1999;   99EP-0402668.
PR   01-MAR-2000; 2000US-0186260.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Denefle P,  Rosier Montus M,  Arnould-Requique I,  Prades C,  Naudin L;
PI   Lemoine C,  Duverger N,  Jaye M,  Searfoss GH,  Romaley A,  Brewer HB;
PI   Dean M;
XX
DR   WPI; 2001-316327/33.
DR   N-PSDB; AAS06135.
XX
PT   New human ABC1 nucleic acids and polypeptides for treating
PT   atherosclerosis, malaria and diabetes -
XX
PS   Claim 46; Page 305 312; 368pp; English.
XX
CC   The sequence represents the amino acid sequence of human ABC1 mutant
CC   polypeptide #11. The polypeptide and nucleic acid sequence, primers and
```

CC probes derived from the ABC1 sequence, and vectors are useful for the
CC prevention of atherosclerosis, in a subject affected by a dysfunction in
CC the reverse transport of cholesterol. The polypeptide encoded by the ABC1
CC gene is useful for screening for an active ingredient for the prevention
CC or treatment of a disease resulting from dysfunction in the reverse
CC transport of cholesterol. The nucleic acids and polypeptides are also
CC useful for treating and preventing cardiovascular and neurological
CC pathologies, and other diseases e.g. Tangier disease, lecithin-
CC cholesterol (LCAT) deficiency, malaria and diabetes
XX
SQ Sequence 1525 AA;

Query Match 92.6%; Score 38; DB 22; Length 1525;
Best Local Similarity 87.5%; Pred. No. 62;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPVE 8
||||:|||
Db 808 nlfespve 815

RESULT 11
AAU02190
ID AAU02190 standard; Protein; 2130 AA.
XX
AC AAU02190;
XX
DT 12-SEP-2001 (first entry)
XX
DE Human ABC1 mutant polypeptide #14.
XX
KW Human; ABC1 gene; atherosclerosis; reverse transport; cholesterol;
KW cardiovascular; neurological; Tangier disease; LCAT deficiency; mutant;
KW lecithin cholesterol acetyltransferase; malaria; diabetes; mutein.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 2112
FT /note= "Frameshift at amino acid 2112 introduces a stop
FT codon at amino acid 2130"
XX
PN WO200130848-A2.
XX
PD 03-MAY-2001
XX
PF 26-OCT-2000; 2000WO-EP10386
XX
PR 26-OCT-1999; 99EP-0402668.
PR 01-MAR-2000; 2000US-0186260.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Denefle P, Rosier-Montus M, Arnould-Requigne I, Prades C, Naudin L;
PI Lemoine C, Duverger N, Jaye M, Scarfoss GH, Pomeley A, Prewer HB;
PI Dean M;
XX
DR WPI; 2001-316327/33.
DR N-PSDB; AAS06138.
XX
PT New human ABC1 nucleic acids and polypeptides for treating
PT atherosclerosis, malaria and diabetes -
XX
PS Claim 46; Page 335-346, 368pp. English.
XX
CC The sequence represents the amino acid sequence of human ABC1 mutant
CC polypeptide #14. The polypeptide and nucleic acid sequence, primers and
CC probes derived from the ABC1 sequence, and vectors are useful for the
CC prevention of atherosclerosis, in a subject affected by a dysfunction in
CC the reverse transport of cholesterol. The polypeptide encoded by the ABC1
CC gene is useful for screening for an active ingredient for the prevention
CC or treatment of a disease resulting from dysfunction in the reverse
CC transport of cholesterol. The nucleic acids and polypeptides are also
CC useful for treating and preventing cardiovascular and neurological
CC pathologies, and other diseases e.g. Tangier disease, lecithin-
CC cholesterol (LCAT) deficiency, malaria and diabetes.
XX
SQ Sequence 2130 AA;

Query Match 92.6%; Score 38; DB 22; Length 2130;
Best Local Similarity 87.5%; Pred. No. 88;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPVE 8
||||:|||
Db 808 nlfespve 815

RESULT 12
AAB38108
ID AAB38108 standard; Protein; 2143 AA.
XX
AC AAB38108;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human ABC1 cholesterol transporter FHA-1 mutant protein (R2144STOP).
XX
KW Human ABC1 cholesterol transporter; chromosome 9q31;
KW ATP binding cassette; HDL deficiency disorder; high density lipoprotein;
KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW cardiovascular disease; coronary artery disease; coronary restenosis;
KW cerebrovascular disease; peripheral vascular disease;
KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW mutein.
XX
OS Homo sapiens.
XX
PN WO200055318-A2.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-IB00532.
XX
PR 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA
PA (XENO-) XENON BIORESEARCH INC.
XX
PI Hayden MR, Wilson AR, Pimstone SN;
XX
DR WPI; 2000-587528/55.
DR N-PSDB; AAC69389.
XX
PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT disease and cancer -
XX
PS Examples; Page -; 229pp; English.
XX
CC The invention relates to the human ABC1 cholesterol transporter protein
CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC a member of the ATP-binding cassette (ABC transporter) superfamily of
CC proteins, and plays a crucial role in cholesterol transport, particularly
CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC located on chromosome 9q31, and mutations in this gene are associated
CC with two genetic HDL (high density lipoprotein) deficiency disorders,
CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases

```
CC   are distinguishable in that TD is an autosomal recessive disorder, while
CC   FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC   cholesterol") in the blood correlate with a high risk of cardiovascular
CC   disease, particularly coronary artery disease, but also cerebrovascular
CC   disease, coronary restenosis, and peripheral vascular disease.
CC   Conversely, a high level of HDL has protective effects against
CC   cardiovascular disease. The invention provides genetic constructs and
CC   transgenic cells and non human animals comprising human ABC1 nucleic
CC   acids, and methods of gene therapy for the treatment or prevention of
CC   cardiovascular disease comprising the administration of an expression
CC   vector encoding ABC1 or an active fragment thereof. The invention also
CC   encompasses compounds which mimic ABC1 activity, compounds which
CC   stimulate ABC1 expression and methods of screening for such compounds.
CC   It further relates to methods for determining whether a patient has an
CC   increased risk for cardiovascular disease due to polymorphisms in the
CC   ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC   or prevent cardiovascular disease, especially coronary artery disease,
CC   cerebrovascular disease, coronary restenosis or peripheral vascular
CC   disease. They may also be used in the treatment of diseases associated
CC   with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC   disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC   The invention specifically excludes proteins with the exact amino acid
CC   sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC   acid with the exact sequence as GenBank Accession No: AI012376.1. The
CC   present sequence represents a mutant human ABC1 cholesterol transporter
CC   associated with an altered cholesterol level and therefore an altered
CC   risk of cardiovascular disease.
CC   Note: The present sequence is not shown in the specification, but is
CC   derived from the native human ABC1 shown on pages 152-157.
XX
SQ   Sequence   2143 AA;


  Query Match             82.6%;  Score 38;  DB 21;  Length 2143;
  Best Local Similarity   87.5%;  Pred. No. 88;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           ||||:|||
Db     808 nlfespve 815


RESULT  13
AAY79380
ID   AAY79380 standard; Protein; 2201 AA.
XX
AC   AAY79380;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Human ATP binding cassette ABCA1 (ABC1) protein.
XX
KW   ABCA1; ABC1; ATP binding cassette; human; cholesterol;
KW   interleukin-1 beta; transporter; inflammation; septic shock;
KW   rheumatoid arthritis; Tangier disease; hypertriglyceridemia;
KW   splenomegaly; atherosclerosis; lipid disorder; dyslipidemia;
KW   psoriasis; lupus erythematosus; diagnosis; therapy.
XX
OS   Homo sapiens.
XX
PN   WO200018912-A2.
XX
PD   06-APR-2000.
XX
PF   21-SEP-1999;   99WO-EP06991.
XX
PR   25-SEP-1998;   98US-0101706.
XX
PA   (FARB ) BAYER AG.
XX
PI   Schmitz G,  Klucken J;
XX
DR   WPI; 2000-293151/25.
DR   N-PSDB; AAZ94734.
XX
FT   Adenosine triphosphate binding proteins useful for identifying agents
FT   for treating atherosclerosis and other inflammatory disorders -
XX
FS   Claim 6; Page 94-105, 154pp; English.
XX
CC   The present sequence is that of human ATP binding cassette protein
CC   ABCA1 (ABC1), the human homologue of mouse ABCA1 (94% identity).
CC   The sequence was deduced from ABCA1 cDNA (see AAZ94734), identified in
CC   a differential screening of human monocytes for cholesterol-sensitive
CC   genes. The ABCA1 gene maps to chromosome 9q22-q31. Dysregulated
CC   ABCA1 is the gene locus involved in the HDL deficiency syndrome
CC   Tangier disease, associated with hypertriglyceridemia and
CC   splenomegaly. ABCA1 is also a transporter for interleukin-1 beta,
CC   making the gene a candidate for treatment of inflammatory diseases
CC   such as rheumatoid arthritis and septic shock. The invention
CC   also provides other cholesterol-sensitive ABC genes (see AAZ94735-63)
CC   that can be used for diagnostic and therapeutic applications,
CC   and for biochemical or cell-based assays to screen for
CC   pharmacologically active compounds useful for the treatment of
CC   lipid disorders, atherosclerosis or other inflammatory diseases
CC   such as psoriasis and lupus erythematosus.
XX
SQ   Sequence   2201 AA;


  Query Match             82.6%;  Score 38;  DB 21;  Length 2201;
  Best Local Similarity   87.5%;  Pred. No. 91;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           ||||:|||
Db     748 nlfespve 755


RESULT  14
AAE13021
ID   AAE13021 standard; Protein; 2201 AA.
XX
AC   AAE13021;
XX
DT   28-JAN-2002  (first entry)
XX
DE   Human ATP binding cassette transporter 1 (ABC1) protein.
XX
KW   Human; ATP binding cassette transporter 1; ABC1; coronary heart disease;
KW   dermatological; atherosclerosis; cardiovascular; inflammatory disease;
KW   psoriasis; lipid disorder; antibacterial; septic shock; gene therapy;
KW   immunosuppressive; lupus erythematosus; rheumatoid arthritis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 978
FT                   /note= "Encoded by CAC"
FT   Misc-difference 980
FT                   /note= "Encoded by AAC"
XX
PN   EP1136552-A1.
XX
PD   26-SEP-2001.
XX
PF   20-MAR-2000; 2000EP-0105820.
XX
PR   20-MAR-2000; 2000EP-0105820.
XX
PA   (FARB ) BAYER AG.
XX
PI   Schmitz G,  Bodzioch M;
XX
DR   WPI; 2001-640388/74.
DR   N-PSDB; AAD21325.
```

```
XX
PT   New adenosine triphosphate binding cassette transporter-1 gene
PT   polymorphisms, useful for diagnosing and treating lipid disorders,
PT   cardiovascular diseases and inflammatory diseases -
XX
PS   Claim 8; Page 23-29; 48pp; English.
XX
CC   The invention relates to four common polymorphisms in the gene encoding
CC   ATP binding cassette transporter-1 (ABC1). ABC1 is associated with
CC   decreased ApoA-1 mediated efflux of cholesterol. The polymorphisms in
CC   ABC1 directly affects cellular lipid homeostasis, which is a key factor
CC   in the atherogenetic processes. The ABC1 polymorphisms are useful for
CC   diagnosing and treating lipid disorders, cardiovascular diseases
CC   (coronary heart disease, atherosclerosis) and inflammatory diseases
CC   (psoriasis, lupus erythematosus). The identification of ABC1 as a
CC   transporter for interleukin-1beta (IL-1beta) identifies this gene as
CC   a candidate for treatment of inflammatory diseases including rheumatoid
CC   arthritis and septic shock. The present sequence is human ABC1 protein
XX
SQ   Sequence   2201 AA;


  Query Match             82.6%;  Score 38;  DB 22;  Length 2201;
  Best Local Similarity   87.5%;  Prod  No  91;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           ||||:|||
Db     748 nlfespve 755


RESULT  15
AAM50227
ID   AAM50227 standard; Protein; 2201 AA.
XX
AC   AAM50227;
XX
DT   07-JAN-2002  (first entry)
XX
DE   Human ATP binding cassette transporter 1 (ABC1).
XX
KW   ATP binding cassette transporter 1, ABC1, human, lipid disorder,
KW   cholesterol, cardiovascular disease, inflammatory disease,
KW   antiinflammatory; antilipaemic; antipsoriatic; dermatological;
KW   Tangier disease; coronary heart disease; diagnosis, gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1136554-A1.
XX
PD   26-SEP-2001.
XX
PF   24-MAR-2000; 2000EP-0106401.
XX
PR   24-MAR-2000; 2000EP-0106401.
XX
PA   (FARB ) BAYER AG.
XX
PI   Schmitz G,  Bodzioch M;
XX
DR   WPI; 2001-640389/74.
DR   N-PSDB; AAI70314.
XX
PT   New adenosine triphosphate binding cassette transporter gene
PT   polymorphisms, useful for diagnosing and treating lipid disorders,
PT   cardiovascular diseases and inflammatory diseases -
XX
PS   Example 1; Page 19-25; 41pp; English.
XX
CC   The present sequence is that of the human adenosine triphosphate
CC   (ATP) binding cassette transporter 1 (ABC1) protein. The invention
CC   provides 4 common polymorphisms in the ABC1 gene (see AAI70314),
CC   all of which alter the amino acid sequence of ABC1 and therefore
CC   may affect its function. The 2 most common polymorphisms are both
CC   associated with a decreased in vitro ApoA-I mediated efflux of
CC   cholesterol from monocyte-derived phagocytes, a feature typical of
CC   Tangier disease. 3 Of the variants are significantly increased in
CC   a population of men having low high density lipoprotein-cholesterol
CC   levels and established coronary heart disease (CHD) relative to
CC   CHD-free control subjects. The use of the provided ABC1 polymorphisms
CC   for the diagnosis and treatment of lipid disorders, cardiovascular
CC   diseases, and inflammatory diseases (e.g. psoriasis, lupus
CC   erythematodes) is claimed. Modulation of ABC1 transcripts or
CC   proteins by antisense or ribozyme technology or DNA decoys is also
CC   claimed.
XX
SQ   Sequence   2201 AA;


  Query Match             82.6%;  Score 38;  DB 22;  Length 2201;
  Best Local Similarity   87.5%;  Prod  No  91;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           ||||:|||
Db     748 nlfespve 755


Search completed: September  5, 2002, 15:26:31
Job time: 371 sec
```

CapCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:28:51 ; Search time 69.78 Seconds
(without alignments)
12.393 Million cell updates/sec

Title: US-09-744-804-38
Perfect score: 46
Sequence: 1 NLFETPVEA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	46	100.0	218	2	A47285	milk fat globule p
2	39	82.6	2201	2	A54774	ATP binding casset
3	37	80.4	381	2	AC2631	conserved hypothet
4	37	80.4	461	2	D97413	hypothetical prote
5	36	78.3	401	2	S65138	glycoprotein anti
6	36	78.3	427	2	S74211	PAS-6/7 protein pr
7	36	78.3	655	2	A10910	probable membrane
8	35	76.1	566	2	JC6058	poly A polymerase
9	35	76.1	658	2	T40107	hypothetical 57.9
10	34	73.9	177	2	T08448	hypothetical prote
11	34	73.9	261	2	S69267	ubiquitous mucin-i
12	34	73.9	409	2	T11743	pP47 protein - pig
13	34	73.9	744	2	T27674	hypothetical prote
14	34	73.9	762	2	T50155	hypothetical prote
15	33	71.7	269	2	B81282	probable periplasm
16	33	71.7	293	2	S04649	beta-lactamase (EC
17	33	71.7	343	2	T26784	hypothetical prote
18	33	71.7	405	2	G84200	cytochrome P450 [i
19	33	71.7	464	2	T16889	hypothetical prote
20	33	71.7	535	1	S76953	protein kinase (EC
21	33	71.7	612	2	T11830	NADH dehydrogenase
22	33	71.7	713	2	G98286	methylmalonyl CoA
23	33	71.7	713	2	AB2997	methylmalonyl CoA
24	33	71.7	714	2	B87543	methylmalonyl CoA
25	32	69.6	179	2	H87418	hypothetical prote
26	32	69.6	196	2	D84074	hypothetical prote
27	32	69.6	233	2	AC3498	SAM-dependent meth
28	32	69.6	291	2	S03230	hypothetical prote
29	32	69.6	312	2	A25455	licheninase (EC 3.
30	32	69.6	329	2	S13735	licheninase (EC 3.
31	32	69.6	334	2	S13734	licheninase (EC 3.
32	32	69.6	334	2	S36235	licheninase (EC 3.
33	32	69.6	334	2	S20026	beta-glucanase - r
34	32	69.6	367	2	A33950	yopM protein - Yer
35	32	69.6	376	2	T19255	[r tein kinase lik
36	32	69.6	389	2	B96806	hypothetical prote
37	32	69.6	400	2	T43599	yop targeted effec
38	32	69.6	418	2	T41087	hypothetical prote
39	32	69.6	422	2	G84702	hypothetical prote
40	32	69.6	504	2	AD3629	vdeC protein [impo
41	32	69.6	571	2	T01511	hypothetical prote
42	32	69.6	744	2	S57061	nuclear pore prote
43	32	69.6	790	2	T22849	hypothetical prote
44	32	69.6	829	2	F89640	protein F59F5.7 [i
45	32	69.6	1269	2	I58375	protein-tyrosine k

ALIGNMENTS

RESULT 1
A47285
milk fat globule protein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 31-Dec 1993 #sequence_revision 31-Dec 1993 #text_change 28 May-1999
C;Accession: A47285
R;Larocca, D.; Peterson, J.A.; Urrea, R.; Kuniyoshi, J.; Bistrain, A.M.; Ceriani, R.L.
Cancer Res. 51, 4994-4998, 1991
A;Title: A Mr 46,000 human milk fat globule protein that is highly expressed in huma
A;Reference number: A47285; MUID:91371351
A;Accession: A47285
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-218 <LAR>
A;Cross-references: GB:S56151; NID:g225396; PIDN:AAB19771.1; PID:g225397
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
F;1-56/Domain: discoidin I amino-terminal homology (fragment) <DN1>
F;60-218/Domain: discoidin I amino-terminal homology <DN2>

Query Match 100.0%; Score 46; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 0.048;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPVEA 9
||||||||| |
Db 25 NLFETPVEA 33

RESULT 2
A54774
ATP binding cassette transporter ABC1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 25-Apr-1995 #sequence_revision 25-Apr-1995 #text_change 02-Feb-2001
C;Accession: A54774
R;Luciani, M.F.; Denizot, F.; Savary, S.; Mattei, M.G.; Chimini, G.
Genomics 21, 150-159, 1994
A;Title: Cloning of two novel ABC transporters mapping on human chromosome 9.
A;Reference number: A54774; MUID:94375008
A;Accession: A54774
A;Molecule type: mRNA
A;Residues: 1-2201 <LUC>
A;Cross-references: GB:X75926; NID:g495256; PIDN:CAA53530.1; PID:g495257
C;Superfamily: unassigned ATP binding cassette proteins; ATP-binding cassette homolog
C;Keywords: ATP; duplication; nucleotide binding; P-loop
F;856-1047/Domain: ATP binding cassette homology <ABC1>
F;873-880/Region: nucleotide-binding motif A (P-loop)
F;1869-2060/Domain: ATP binding cassette homology <ABC2>
F;1886-1893/Region: nucleotide-binding motif A (P-loop)

Query Match 82.6%; Score 38; DB 2; Length 2201;

Best Local Similarity 87.5%; Pred. No. 32;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NLFETPVE 8
            ||||:|||
Db     748 NLFESPVE 755
```

RESULT 3
AC2631
conserved hypothetical protein Atu0446 [imported] - Agrobacterium tumefaciens (strain C5
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AC2631
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, S.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClel
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AC2631
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-381 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL41465.1; PID:g17738790; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu0446
A;Map position: circular chromosome

Query Match 80.4%; Score 37; DB 2; Length 381;
Best Local Similarity 77.8%; Pred. No. 7.2;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 NLFETPVEA 9
            ||:||| ||
Db     180 NLYETPPEA 188
```

RESULT 4
D97413
hypothetical protein AGR_C_790 [imported] - Agrobacterium tumefaciens (strain C58, Dared
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: D97413
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: D97413
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-461 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK88701.1; PID:g15155364; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_790
A;Map position: circular chromosome

Query Match 80.4%; Score 37; DB 2; Length 461;
Best Local Similarity 77.8%; Pred. No. 8.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 NLFETPVEA 9
            ||.||| ||
Db     260 NLYETPPEA 268
```

RESULT 5
S65138
glycoprotein antigen MGP57/53, mammary gland - bovine (fragment)
N;Alternate names: glycoprotein component 2/major fat-globule membrane protein/MFG-E
C;Species: Bos primigenius taurus (cattle)
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
C;Accession: S65138; G48394
R;Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A;Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal
A;Reference number: S65138; MUID:96125746
A;Accession: S65138
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-401 <AOK>
R;Mather, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A;Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-
II-like sequences.
A;Reference number: A48394; MUID:93250576
A;Accession: G48394
A;Status: preliminary
A;Molecule type: protein
A;Residues: 207-220 <MAT>
A;Experimental source: milk
A;Note: sequence extracted from NCBI backbone (NCBIP:131457)
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
C;Keywords: glycoprotein
F;1-32/Domain: EGF homology (fragment) <EG1>
F;40-79/Domain: EGF homology <EG2>
F;82-233/Domain: discoidin I amino-terminal homology <DN1>
F;243-401/Domain: discoidin I amino terminal homology <DN2>

Query Match 78.3%; Score 36; DB 2; Length 401;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NLFETPVE 8
            |||:|||:|
Db     208 NLFDTPLE 215
```

RESULT 6
S74211
PAS 6/7 protein precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000
C;Accession: S74211; S78114; S24181
R;Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
Eur. J. Biochem. 240, 628-636, 1996
A;Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat g
A;Reference number: S74211; MUID:97008954
A;Accession: S74211
A;Molecule type: mRNA
A;Residues: 1-427 <HVA>
A;Cross-references: EMBL:X91809; NID:g1542778; PIDN:CAA62997.1; PID:g1542779
A;Accession: S78114
A;Molecule type: protein
A;Residues: 19-85,96-110,140-165,174-226,221-232,248-277,285-293,309-337,396-420,425-
R;Kim, D.H.; Kanno, C.; Mizokami, Y.
Biochim. Biophys. Acta 1122, 203-211, 1992
A;Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, f
A;Reference number: S23926; MUID:92353107
A;Accession: S24181
A;Molecule type: protein
A;Residues: 383-394 <KIM>
C;Superfamily: milk fat globule protein; discoidin I amino terminal homology; EGF hom
C;Keywords: blocked amino end; disulfide bond; glycoprotein; milk
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-427/Product: PAS-6/7 protein #status experimental <MAT>
F;24-58/Domain: EGF homology <EG1>
F;66-105/Domain: EGF homology <EG2>
F;108-265/Domain: discoidin I amino-terminal homology <DN1>

```
F;269-427/Domain. disulfide-terminal homology <TN>
F;21-35,29-42,44-58,66-77,71-94,96-105/Disulfide bonds #status predicted
F;27/Binding site: carbohydrate (Ser) (covalent) #status experimental
F;34/Binding site: carbohydrate (Thr) (covalent) #status experimental
F;55,227/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;109-165,252-256,170-427/Disulfide bonds #status experimental

  Query Match             78.3%;  Score 36;  DB 2;  Length 427;
  Best Local Similarity   75.0%;  Pred. No. 13;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 NLFETPVE 8
              |||:|||:|
Db        234 NLFETTLE 241

RESULT   7
AI0910
probable membrane protein STY3545 [imported] - Salmonella enterica subsp. enterica serov
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C;Accession: AI0910
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0592; PMID:11677608
A;Accession: AI0910
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-655 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD07879.1; PID:g16503126; GSPDB:GN00172
C;Genetics:
A;Gene: STY3545

  Query Match             78.3%;  Score 36;  DB 2;  Length 655;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LFETPVE 8
              |||||||
Db        273 LFETPVE 279

RESULT   8
JC6058
poly A polymerase (EC 2.7.7.-) - fission yeast (Schizosaccharomyces pombe)
N;Alternate names: PLA1 protein
C;Species: Schizosaccharomyces pombe
C;Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 20-Jun-2000
C;Accession: JC6058; T40580
R;Ohnacker, M.; Minvielle-Sebastia, L.; Keller, W.
Nucleic Acids Res. 24, 2585-2591, 1996
A;Title: The Schizosaccharomyces pombe pla1 gene encodes a poly(A) polymerase and can f
A;Reference number: JC6058; MUID:96293264
A;Accession: JC6058
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-566 <OHN>
A;Cross-references: EMBL:X73376; NID:g1419055; PIDN:CAA56141.1; PID:g1419056
A;Experimental source: strain JL17-3A
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A;Reference number: Z21938
A;Accession: T40580
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-566 <SEE>
A;Cross-references: EMBL:AL035116; PIDN:CAA22884.1; GSPDB:GN00067; SPDB:SPBC646.04
A;Experimental source: strain 972h-; cosmid c646
C;Comment: This enzyme catalyzes unspecific polyadenylation.
C;Genetics:
A;Gene: pla1; SPBC646.04
A;Map position: 2
A;Introns: 486/1; 525/2; 556/3
C;Keywords: nucleotidyltransferase; RNA binding
F;45-165/Domain: catalytic core #status predicted <CCA>
F;537-547/Domain: RNA-binding #status predicted <RNB>
F;99,101,153/Active site: Asp #status predicted

  Query Match             76.1%;  Score 35;  DB 2;  Length 566;
  Best Local Similarity   66.7%;  Pred. No. 30;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 NLFETPVEA 9
              ||||:| |:
Db         36 NLFESPAES 44

RESULT   9
T40107
hypothetical 57.9 kd protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 04-Dec-1998 #sequence_revision 04-Dec-1998 #text_change 04-Dec-1999
C;Accession: T40107
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Taylor, K.; Harris, D.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21906
A;Accession: T40107
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-658 <WOO>
A;Cross-references: EMBL:AL031786; PIDN:CAA21166.1; GSPDB:GN00067; SPDB:SPBC2D10.04
A;Experimental source: strain 972h-; cosmid c2D10
C;Genetics:
A;Gene: SPDB:SPBC2D10.04
A;Map position: 2

  Query Match             76.1%;  Score 35;  DB 2;  Length 658;
  Best Local Similarity   87.5%;  Pred. No. 35;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 NLFETPVE 8
              |||||| |
Db        423 NLFETPDE 430

RESULT  10
T08448
hypothetical protein F2206.110 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C;Accession: T08448
R;Quetier, F.; Parmello, B.; Boutry, M.; Goffeau, A.; Salanoubat, M.; Mewes, H.W.; Ma
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16420
A;Accession: T08448
A;Molecule type: DNA
A;Residues: 1-177 <QUE>
A;Cross-references: EMBL:AL050300; GSPDB:GN00061; ATSP:F2206.110
A;Experimental source: cultivar Columbia; BAC clone F2206
C;Genetics:
A;Gene: ATSP:F2206.110
A;Map position: 3

  Query Match             73.9%;  Score 34;  DB 2;  Length 177;
  Best Local Similarity   77.8%;  Pred. No. 13;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy       1 NLFETPVEA 9
           || | ||||
Db      66 NLIELPVEA 74


RESULT  11
D69267
ubiquinone/menaquinone biosynthesis methyltransferase (ubiE) homolog - Archaeoglobus ful
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C;Accession: D69267
R;Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, J.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate reducing archaeo
A;Reference number: A69250; MUID:98049343
A;Accession: D69267
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-261 <KLE>
A;Cross-references: GB:AE000782; NID:g2384419; PIDN:AAB91083.1; PID:g2
C;Superfamily: bioC homology
F;61-159/Domain: bioC homology <BIOC>


  Query Match             73.9%;  Score 34;  DB 2;  Length 261;
  Best Local Similarity   75.0%;  Pred. No. 20;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           :||||||:|
Db      35 DLFETPME 42


RESULT  12
T11743
pP47 protein - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T11743
R;Ensslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toepfe
Biol. Reprod. 58, 1057-1064, 1998
A;Title: Molecular cloning and characterization of P47, a novel boar sperm-associated ac
A;Reference number: Z17325; MUID:98206817
A;Accession: T11743
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-409 <ENS>
A;Cross-references: EMBL:Y11683; NID:g2652927; PIDN:CAA72379.1; PID:g2652928
A;Experimental source: testis
C;Function:
A;Description: may be involved in membrane remodeling and/or function as a zona pellucid
C;Superfamily: unassigned EGF-related proteins; EGF homology
F;6-40/Domain: EGF homology <EGF>


  Query Match             73.9%;  Score 34;  DB 2;  Length 409;
  Best Local Similarity   75.0%;  Pred. No. 33;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           |||| |:|
Db     216 NLFEVPLE 223


RESULT  13
T27674
hypothetical protein ZK1058.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T27674
R;Mortimore, B.
submitted to the EMBL Data Library, August 1994
A;Reference number: Z20403
A;Accession: T27674
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-744 <WIL>
A;Cross-references: EMBL:Z35604; PIDN:CAA84676.1; GSPDB:GN00021; CESP:ZK1058.1
A;Experimental source: clone ZK1058
C;Genetics:
A;Gene: CESP:ZK1058.1
A;Map position: 3
A;Introns: 137/2; 353/3; 439/3; 530/3; 695/2
C;Superfamily: human methylmalonyl-CoA mutase; cobalamin-binding homology


  Query Match             73.9%;  Score 34;  DB 2;  Length 744;
  Best Local Similarity   75.0%;  Pred. No. 65;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LFETPVEA 9
           |:|||:||
Db     541 LFSIPFEA 548


RESULT  14
T50155
hypothetical protein SPAC222.14c [imported] - fission yeast (Schizosaccharomyces pomb
C;Species: Schizosaccharomyces pombe
C;Date: 04-Jun-2000 #sequence_revision 04-Jun-2000 #text_change 21-Jul-2000
C;Accession: T50155
R;Hamlin, N.; Churcher, C.M.; McDougall, K.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, November 1999
A;Reference number: Z25043
A;Accession: T50155
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-762 <RAM>
A;Cross-references: EMBL:AL132798; PIDN:CAB67706.1; GSPDB:GN00062; SPDB:SPAC222.14c
A;Experimental source: strain 972h(-); cosmid c222
C;Genetics:
A;Gene: SPDB:SPAC222.14c
A;Map position: 1
A;Introns: 53/1; 78/1; 106/3
C;Superfamily: Saccharomyces cerevisiae probable membrane protein YOR165w


  Query Match             73.9%;  Score 34;  DB 2;  Length 762;
  Best Local Similarity   66.7%;  Pred. No. 67;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |:||| :||
Db     328 NMFETVIEA 336


RESULT  15
E81282
probable periplasmic protein Cj1376 [imported] - Campylobacter jejuni (strain NCTC 11
C;Species: Campylobacter jejuni
C;Date: 01-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C;Accession: E81282
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A;Reference number: A81250; MUID:20150912
A;Accession: E81282
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-269 <PAR>
```

A;Cross-references: GB:AL149078; GB:AL111168; NID:g6968723; PIDN:CAB73803.1; PID:g6968811
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj1376

Query Match 71.7%; Score 33; DB 2; Length 269;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 NLFETPVE 8
             ||| |||:
Db      80 NLFNTPVK 87
```

Search completed: September 5, 2002, 15:28:52
Job time: 457 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:41 ; Search time 33.99 Seconds
(without alignments)
10.252 Million cell updates/sec

Title: US-09-744-804-38
Perfect score: 46
Sequence: 1 NLPETPVEA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40 *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	46	100.0	387	1	MFGM_HUMAN	Q08431 homo sapien
2	38	82.6	2261	1	ABC1_HUMAN	O95477 homo sapien
3	38	82.6	2261	1	ABC1_MOUSE	P41233 mus musculu
4	36	78.3	427	1	MFGM_BOVIN	Q95114 bos taurus
5	34	73.9	409	1	MFGM_PIG	P79385 sus scrofa
6	34	73.9	744	1	MUTA_CAEEL	Q23381 caenorhabdi
7	33	71.7	293	1	BLAC_RHOCA	P14171 rhodobacter
8	33	71.7	464	1	YSV5_CAEEL	Q10011 caenorhabdi
9	33	71.7	535	1	SPKC_SYNY3	P74745 synechocyst
10	33	71.7	612	1	NU5M_GADMO	P55782 gadus morhu
11	32	69.6	291	1	Y32K_SSV1	P20197 sulfolobus
12	32	69.6	312	1	GUB2_HORVU	P12257 hordeum vul
13	32	69.6	409	1	YOPM_YERPE	P17778 yersinia pe
14	32	69.6	744	1	NUB5_YPAST	P46673 saccharomyc
15	32	69.6	1363	1	VCR3_MOUSE	P35317 mus musculu
16	31	67.4	62	1	RS21_ANASP	P49224 anabaena sp
17	31	67.4	217	1	GEE1_HUMAN	Q9hav7 homo sapien
18	31	67.4	217	1	GPE1_PAT	P97576 rattus norv
19	31	67.4	255	1	Y888_HELPJ	Q9zkw3 helicobacte
20	31	67.4	255	1	Y888_HELPY	O25732 helicobacte
21	31	67.4	290	1	MFPC_BACSU	Q01466 bacillus su
22	31	67.4	334	1	PEUR_BACSU	P40410 bacillus su
23	31	67.4	475	1	YDF6_SCHPO	Q10478 schizosacch
24	31	67.4	492	1	GATB_ARCFU	O28164 archaeoglob
25	31	67.4	581	1	LEPA_MYCPN	P75498 mycoplasma
26	31	67.4	598	1	LEPA_MYCGE	P47384 mycoplasma
27	31	67.4	665	1	SIN1_SCHPO	Q9p7y9 schizosacch
28	31	67.4	712	1	MUTB_RHIME	O86828 rhizobium m
29	31	67.4	715	1	MUTB_PORGI	Q59677 porphyromon
30	31	67.4	825	1	TI4R_HUMAN	P24394 homo sapien
31	31	67.4	1385	1	FAT1_SCHPO	O14245 schizosacch
32	30	65.2	175	1	Y397_ARCFU	O28860 archaeoglob
33	30	65.2	190	1	NUDH_XYLFA	Q9pga9 xylella fas
34	30	65.2	231	1	ALSE_ECOLI	P32719 escherichia
35	30	65.2	246	1	PI4_HALMA	P12735 halobacula
36	30	65.2	261	1	YHB6_SYNY3	P73335 synechocyst
37	30	65.2	345	1	KR1_SVVD	Q04543 simian vari
38	30	65.2	346	1	DHAS_HELPI	Q9zk28 helicobacte
39	30	65.2	346	1	DHAS_HELPY	O25801 helicobacte
40	30	65.2	366	1	P53_PLAFE	O12946 platichthys
41	30	65.2	388	1	NCPA_ASPFA	Q00258 aspergillus
42	30	65.2	469	1	KCC4_MOUSE	P08414 mus musculu
43	30	65.2	474	1	F504_HUMAN	Q15056 homo sapien
44	30	65.2	474	1	KCC4_RAT	P13234 rattus norv
45	30	65.2	498	1	YAC3_SCHPO	Q09817 schizosacch

ALIGNMENTS

RESULT 1
MFGM_HUMAN
ID MFGM_HUMAN STANDARD; PRT; 387 AA.
AC Q08431;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule EGF factor 8) (MFG-E8) (HMFG)
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
GN MFGE8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast, and Breast carcinoma;
RX MEDLINE=96213908; PubMed=8639264;
RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
RT "Cloning and sequence analysis of human breast epithelial antigen
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal
RT growth factor-like domain.";
RL DNA Cell Biol. 15:281-286(1996).
RN [2]
RP SEQUENCE OF 170-387 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=91371351; PubMed=1909932;
RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
RA Ceriani R.L.;
RL "A Mr 46,000 human milk fat globule protein that is highly expressed
RT in human breast tumors contains factor VIII-like domains.";
RL Cancer Res. 51:4994-4998(1991).
RN [3]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=98194924; PubMed=9535276;
RA Giuffrida M.G., Cavaletto M., Giunta C., Conti A.,
RA Godovac-Zimmermann J.;
RT "Isolation and characterization of full and truncated forms of human
RT breast carcinoma protein BA46 from human milk fat globule membranes.";
RL J. Protein Chem. 17:143-148(1998).
RN [4]
RP SEQUENCE OF 1-387, AND IDENTIFICATION OF MEDIN.
RX MEDLINE=99342076; PubMed=10411933;
RA Hanggvist B., Nansland T., Sletten K., Westermark G.T., Mucchiano G.,
RA Tjernberg L.O., Nordstedt C., Engstrom U., Westermark P.;
RL "Medin: an integral fragment of aortic smooth muscle cell-produced
RT lactadherin forms the most common human amyloid.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:8074-8074(1999).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=97405885; PubMed=9260929;
RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
RT (RGD)-dependent cell adhesion.";

```
RL   DNA Cell Biol. 16:861-869(1997).
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
CC       SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
CC   -!- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
CC       MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC   -!- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
CC       AMINO ACID 264 AND 273.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U58516; AAC50549.1; -.
DR   EMBL; S56151; AAB19771.1; -.
DR   MIM; 602281; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid
FT   SIGNAL       1     23
FT   CHAIN       24    387       LACTADHERIN.
FT   CHAIN      202    387       LACTADHERIN, SHORT FORM
FT   CHAIN      268    317       MEDIN.
FT   DOMAIN      24     67       EGF-LIKE.
FT   DOMAIN      70    225       F5/8 TYPE C 1.
FT   DOMAIN     230    387       F5/8 TYPE C 2.
FT   SITE        46     48       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    27     38       BY SIMILARITY.
FT   DISULFID    32     55       BY SIMILARITY.
FT   DISULFID    57     66       BY SIMILARITY.
FT   DISULFID    70    225       BY SIMILARITY.
FT   DISULFID   212    216       BY SIMILARITY.
FT   DISULFID   230    387       BY SIMILARITY
FT   CARBOHYD   238    238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   325    325       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   329    329       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   350    350       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   387 AA;  43123 MW;  2EE6571DEC83782D CRC64,


  Query Match          100.0%;  Score 46;  DB 1;  Length 387;
  Best Local Similarity  100.0%;  Pred. No. 0.033;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |||||||||
Db     194 NLFETPVEA 202


RESULT   2
ABC1_HUMAN
ID   ABC1_HUMAN      STANDARD;      PRT;  2261 AA.
AC   O95477; Q9UN08; Q9UN07; Q9UN06; Q9NQV4; Q9UN04; Q96P85; Q96S56;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette
DE   transporter 1) (ATP-binding cassette 1) (ABC 1) (Cholesterol efflux
DE   regulatory protein).
GN   ABCA1 OR ABC1 OR CERP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20345099; PubMed=10884428;
RA   Santamarina-Fojo S., Peterson K.M., Knapper C.L., Qiu Y.,
RA   Freeman L.A., Cheng J.-F., Osorio J., Remaley A.T., Yang X.-P.,
RA   Haudenschild C.C., Prades C., Chimini G., Blackmon E.E.,
RA   Francois T.L., Duverger N., Rubin E.M., Rosier M., Denefle P.,
RA   Fredrickson D.S., Brewer H.B. Jr.;
RT   "Complete genomic sequence of the human ABCA1 gene: analysis of the
RT   human and mouse ATP binding cassette A promoter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:7987-7992(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Schwartz K., Lawn R.M., Wade D.P.;
RT   "ABCA1 gene expression and apoA-I-mediated cholesterol efflux are
RT   regulated by LXR.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21251004; PubMed=11352567;
RA   Qiu Y., Cavelier L., Chiu S., Yang X., Rubin E., Cheng J.-F.;
RT   "Human and mouse ABCA1 comparative sequencing and transgenesis
RT   studies revealing novel regulatory sequences.";
RL   Genomics 73:66-76(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Tanaka A.R., Abe-Dohmae S., Arakawa R., Sadahami K., Kidera A.,
RA   Kioka N., Amachi T., Yokoyama S., Ueda K.;
RT   "A new topological model of functional human ABCA1 signal peptide
RT   cleavage and glycosylation of a large extracellular domain.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 21-2261 FROM N.A.
RX   MEDLINE=99194549; PubMed=10092505;
RA   Langmann T., Klucken J., Reil M., Liebisch G., Luciani M.F.,
RA   Chimini G., Kaminski W.E., Schmitz G.;
RT   "Molecular cloning of the human ATP-binding cassette transporter 1
RT   (hABC1): evidence for sterol-dependent regulation in macrophages.";
RL   Biochem. Biophys. Res. Commun. 257:29-33(1999).
RN   [6]
RP   SEQUENCE OF 21-2261 FROM N.A.
RX   MEDLINE=99364413; PubMed=10431238;
RA   Rust S., Rosier M., Funke H., Real J., Amoura Z., Piette J.-C.,
RA   Deleuze J.-F., Brewer H.B., Duverger N., Denefle P., Assmann G.;
RT   "Tangier disease is caused by mutations in the gene encoding
RT   ATP-binding cassette transporter 1.";
RL   Nat. Genet. 22:352-355(1999).
RN   [7]
RP   VARIANTS FHA THR-1091 AND 1893-GLU-ASP-1894 DEL.
RX   MEDLINE=20001430; PubMed=10533863;
RA   Marcil M., Brooks-Wilson A., Clee S.M., Roomp K., Zhang L.H., Yu L.,
RA   Collins J.A., van Dam M., Molhuizen H.O.F., Loubser O.,
RA   Ouelette B.F.F., Sensen C.W., Fichter K., Mott S., Denis M.,
RA   Boucher B., Pimstone S., Genest J. Jr., Kastelein J.J.P., Hayden M.R.;
RT   "Mutations in the ABC1 gene in familial HDL deficiency with defective
RT   cholesterol efflux.";
RL   Lancet 354:1341-1346(1999).
RN   [8]
RP   VARIANTS TD ARG-597 AND ARG-1477, AND VARIANT FHA LEU-693 DEL.
RX   MEDLINE=99364411; PubMed=10431236;
RA   Brooks-Wilson A., Marcil M., Clee S.M., Zhang L.-H., Roomp K.,
RA   van Dam M., Yu L., Brewer C., Collins J.A., Molhuizen H.O.F.,
RA   Loubser O., Ouelette B.F.F., Fichter K., Ashbourne-Excoffon K.J.D.,
RA   Sensen C.W., Scherer S., Mott S., Denis M., Martindale D.,
RA   Frohlich J., Morgan K., Koop B., Pimstone S., Kastelein J.J.P.,
RA   Hayden M.R.;
```

RT "Mutations in ABC1 in Tangier disease and familial high-density
RT lipoprotein deficiency.";
RL Nat. Genet. 22:336-345(1999).
RN [9]
RP VARIANTS TD SER-590; SER-935 AND VAL-937, AND VARIANTS ALA-399 AND
RP MET-883.
RX MEDLINE=99364413; PubMed=10431237;
RA Bodzioch M., Orso E., Klucken J., Langmann T., Boettcher A.,
RA Diederich W., Drobnik W., Barlage S., Buechler C.,
RA Porsch-Oezcueruemez M., Kaminski W.E., Hahmann H.W., Oette K.,
RA Rothe G., Aslanidis C., Lackner K.J., Schmitz G.;
RT "The gene encoding ATP-binding cassette transporter 1 is mutated in
RT Tangier disease.";
RL Nat. Genet. 22:347-351(1999).
RN [10]
RP VARIANTS TD ILE-929; ARG-597 AND ARG-1477, AND VARIANTS FHA LEU 693
RP DEL; THR-1091-ARG-GLN-ASP-1094 DEL AND LEU-2150.
RX MEDLINE=20540002; PubMed=11086027;
RA Clee S.M., Kastelein J.J.P., van Dam M., Marcil M., Roomp K.,
RA Zwarts K.Y., Collins J.A., Roelants R., Yamasawa N., Stulc T.,
RA Suda T., Ceska R., Boucher B., Rondeau C., DeSouich C.,
RA Brooks-Wilson A., Molhuizen H.O.F., Frohlich J., Genest J. Jr.,
RA Hayden M.R.;
RT "Age and residual cholesterol efflux affect HDL cholesterol levels and
RT coronary artery disease in ABCA1 heterozygotes.";
RL J. Clin. Invest. 106:1263-1270(2000).
RN [11]
RP VARIANTS TD ASN-1289 AND HIS-1800.
RX MEDLINE=20171564; PubMed=10706591;
RA Brousseau M.E., Schaefer E.J., Dupuis J., Eustace B.,
RA Van Eerdewegh P., Goldkamp A.L., Thurston L.M., FitzGerald M.G.,
RA Yasek-McKenna D., O'Neill G., Eberhart G.P., Wolfenbach R.,
RA Ordovas J.M., Freeman M.W., Brown R.H. Jr., Gu J.Z.;
RT "Novel mutations in the gene encoding ATP-binding cassette 1 in four
RT tangier disease kindreds.";
RL J. Lipid Res. 41:453-441(2000).
RN [12]
RP VARIANT TD ASP-1045, VARIANT FHA CYS-230, AND VARIANTS LYS-219,
RP ILE-825; MET-883 AND LYS-1587.
RX MEDLINE=20396633; PubMed=10938021;
RA Wang J., Burnett J.R., Near S., Young K., Zinman B., Hanley A.J.G.,
RA Connelly P.W., Harris S.B., Hegele R.A.;
RT "Common and rare ABCA1 variants affecting plasma HDL cholesterol.";
RL Arterioscler. Thromb. Vasc. Biol. 20:1983-1989(2000).
RN [13]
RP VARIANT TD TRP-587, AND VARIANT LEU-2168.
RX MEDLINE=21157002; PubMed=11257260;
RA Bertolini S., Pisciotta L., Seri M., Cusano R., Cantafora A.,
RA Calabresi L., Franceschini G., Ravazzolo R., Calandra S.,
RT "A point mutation in ABC1 gene in a patient with severe premature
RT coronary heart disease and mild clinical phenotype of Tangier
RT disease.";
RL Atherosclerosis 154:599-605(2001).
RN [14]
RP VARIANTS LYS-219; MET-883 AND ASP-1172.
RX MEDLINE=21157003; PubMed=11257261;
RA Brousseau M.E., Bodzioch M., Schaefer E.J., Goldkamp A.L., Kielar D.,
RA Probst M., Ordovas J.M., Aslanidis C., Lackner K.J.,
RA Bloomfield Rubins H., Collins D., Robins S.J., Wilson P.W.F.,
RA Schmitz G.;
RT "Common variants in the gene encoding ATP-binding cassette transporter
RT 1 in men with low HDL cholesterol levels and coronary heart disease.";
RL Atherosclerosis 154:607-611(2001).
RN [15]
RP VARIANT TD LEU-1506.
RX MEDLINE=21369429; PubMed=11470961;
RA Lapicka-Bodzioch K., Bodzioch M., Kraeli M., Kielar D., Probst M.,
RA Kiec B., Andrikovics H., Boettcher A., Rubacek J., Aslanidis C.,
RA Suttorp N., Schmitz G.;
RT "Homogeneous assay based on 52 primer sets to scan for mutations of
RT the ABCA1 gene and its application in genetic analysis of a new
RT patient with familial high-density lipoprotein deficiency syndrome.";
RL Biochim. Biophys. Acta 1537:42-48(2001).

RN [16]
RP VARIANTS TD ASN 1289 AND HIS-1800, AND VARIANT LYS 219.
RX MEDLINE=21369433; PubMed=11470965;
RA Huang W., Moriyama K., Koga T., Hua H., Ageta M., Kawabata S.,
RA Mawatari K., Imamura T., Eto T., Kawamura M., Teramoto T., Sasaki J.;
RT "Novel mutations in ABCA1 gene in Japanese patients with Tangier
RT disease and familial high density lipoprotein deficiency with
RT coronary heart disease.";
RL Biochim. Biophys. Acta 1537:71-78(2001).
RN [17]
RP VARIANTS LYS-219; ALA-399; MET-771; PRO-774; ASN-776; ILE-825;
RP MET-883; ASP-1172; LYS-1587 AND CYS-1731.
RX MEDLINE=21128379; PubMed=11238261;
RA Clee S.M., Zwinderman A.H., Engert J.C., Zwarts K.Y.,
RA Molhuizen H.O.F., Roomp K., Jukema J.W., van Wijland M., van Dam M.,
RA Hudson T.J., Brooks Wilson A., Genest J. Jr., Kastelein J.J.P.,
RA Hayden M.R.;
RT "Common genetic variation in ABCA1 is associated with altered
RT lipoprotein levels and a modified risk for coronary artery disease.";
RL Circulation 103:1198-1205(2001).
RN [18]
RP VARIANT TD THR-255, AND VARIANT ATHEROSCLEROSIS ASP-1611.
RX MEDLINE=21645894; PubMed=11785958;
RA Nishida Y., Hirano K., Tsukamoto K., Nagano M., Ikegami C.,
RA Ishihara M., Sasaki N., Zhang J., Isaji K., Matsuyama A., Adachi T.,
RA Matsuura F., Ishigami M., Sakai N., Hiraoka H., Hattori H.,
RA Wellington C., Yoshida Y., Misugi S., Hayden M.R., Egashira T.,
RA Yamashita S., Matsuzawa Y.;
RT "Expression and functional analyses of novel mutations of ATP-binding
RT cassette transporter 1 in Japanese patients with high-density
RT lipoprotein deficiency.";
RL Biochem. Biophys. Res. Commun. 290:714-721(2002).
CC -!- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION
CC TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL
CC TRANSPORT.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED, BUT MOST ABUNDANT IN
CC MACROPHAGES.
CC -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.
CC -!- DISEASE: DEFECTS IN ABCA1 ARE A CAUSE OF HIGH DENSITY LIPOPROTEIN
CC DEFICIENCY TYPE I (HDLDI), ALSO KNOWN AS TANGIER DISEASE (TD). TD
CC IS A RECESSIVE DISORDER CHARACTERIZED BY ABSENCE OF HIGH DENSITY
CC LIPOPROTEIN (HDL) CHOLESTEROL FROM PLASMA, HEPATOSPLENOMEGALY,
CC PERIPHERAL NEUROPATHY, AND FREQUENTLY PREMATURE CORONARY ARTERY
CC DISEASE (CAD).
CC -!- DISEASE: Defects in ABCA1 are a cause of high density lipoprotein

Query Match 32.6%; Score 38; DB 1; Length 2261;
Best Local Similarity 87.5%; Pred. No. 10;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPVE 8
||||:|||
Db 808 NLFESPVE 815

RESULT 3
ABC1_MOUSE
ID ABC1_MOUSE STANDARD; PRT; 2261 AA.
AC P41233;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub family A, member 1 (ATP binding cassette
DE transporter 1) (ATP-binding cassette 1) (ABC-1).
GN ABCA1 OR ABC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```
RC   STRAIN=DBA/2; TISSUE=Macrophage;
RX   MEDLINE=94375008; PubMed=8088782;
RA   Luciani M.F., Denizot F., Savary S., Mattei M.-G., Chimini G.;
RT   "Cloning of two novel ABC transporters mapping on human chromosome
RT   9.";
RL   Genomics 21:150-159(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RA   Qiu Y., Cavelier L., Chiu S., Rubin E., Cheng J.-F.;
RT   "Human and mouse ABCA1 comparative sequencing and transgenesis studies
RT   identify potential regulatory sequences.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION
CC       TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL
CC       TRANSPORT (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED IN ADULT TISSUES. HIGHEST
CC       LEVELS ARE FOUND IN PREGNANT UTERUS AND UTERUS.
CC   -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC       EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC       ATP BINDING CASSETTE (ABC) DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X75926; CAA53530.1; ALT_INIT.
DR   EMBL; AF287263; AAG39073.1; ALT_INIT.
DR   MGD; MGI:99607; Abca1.
DR   InterPro; IPR003593; AAA
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A
DR   InterPro; IPR003838; DUF214.
DR   InterPro; IPR000897; SRP54
DR   Pfam; PF00005; ABC_tran; 2.
DR   Pfam; PF02687; DUF214; 1.
DR   Pfam; PF00448; SRP54; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1
KW   ATP binding; Glycoprotein; Transmembrane; Transport.
FT   TRANSMEM     26     42       POTENTIAL.
FT   TRANSMEM    640    656       POTENTIAL.
FT   TRANSMEM    690    706       POTENTIAL.
FT   TRANSMEM    717    733       POTENTIAL.
FT   TRANSMEM    749    765       POTENTIAL.
FT   TRANSMEM    771    787       POTENTIAL.
FT   TRANSMEM   1041   1057       POTENTIAL.
FT   TRANSMEM   1351   1367       POTENTIAL.
FT   TRANSMEM   1661   1677       POTENTIAL.
FT   TRANSMEM   1708   1724       POTENTIAL.
FT   TRANSMEM   1737   1753       POTENTIAL.
FT   TRANSMEM   1775   1791       POTENTIAL.
FT   TRANSMEM   1854   1870       POTENTIAL.
FT   NP_BIND     933    940       ATP (POTENTIAL).
FT   NP_BIND    1946   1953       ATP (POTENTIAL).
FT   CARBOHYD     14     14       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    244    244       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    337    337       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    349    349       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400    400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    489    489       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    521    521       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    829    829       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1144   1144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1294   1294       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1453   1453       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1499   1499       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1504   1504       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1637   1637       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2044   2044       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2238   2238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1567   1568       MISSING (IN REF. 2)
FT   CONFLICT   2024   2024       MISSING (IN REF. 2)
SQ   SEQUENCE   2261 AA;  254011 MW;  FAE62B2FFD1D05F9 CRC64;

  Query Match             82.6%;  Score 95;  DB 2;  Length 2261;
  Best Local Similarity   87.5%;  Pred. No. 10;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           ||||:|||
Db     808 NLFESPVE 815


RESULT   4
MFGM_BOVIN
ID   MFGM_BOVIN     STANDARD;      PRT;   427 AA.
AC   Q95114; Q27959; P79344;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
DE   (MGF57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
DE   SP47) (BP47) (Components 15/16).
GN   MFGE8.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC   STRAIN=HOLSTEIN; TISSUE=Mammary gland;
RX   MEDLINE=97008954; PubMed=8856064;
RA   Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RA   Petersen T.E.;
RT   "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT   milk fat globules.";
RL   Eur. J. Biochem. 240:628-636(1996).
RN   [2]
RP   SEQUENCE OF 18-427 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=96125736; PubMed=8541316;
RA   Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA   Matsuda T.;
RT   "Molecular cloning of glycoprotein antigens MGP57/53 recognized by
RT   monoclonal antibodies raised against bovine milk fat globule
RT   membrane.";
RL   Biochim. Biophys. Acta 1245:385-391(1995).
RN   [3]
RP   SEQUENCE OF 19-427 FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 140-146; 174-187; 244-246 AND 422-427
RC   TISSUE=Milk;
RX   MEDLINE=93250576; PubMed=8485470;
RA   Mather I.H., Banghart L.R., Lane W.S.;
RT   "The major fat-globule membrane proteins, bovine components 15/16 and
RT   guinea-pig GP 55, are homologous to MFG-E8, a murine glycoprotein
RT   containing epidermal growth factor-like and factor V/VIII-like
RT   sequences.";
RL   Biochem. Mol. Biol. Int. 29:545-554(1993).
```

```
CC   -!- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
CC       MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM
CC       LACKS 53 AMINO ACIDS WITHIN THE F5/8 TYPE C 1 DOMAIN.
CC   -!- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN
CC   -!- PTM: THE 2 O-LINKED GLYCANS CONSIST OF GAL, GLCNAC AND FUC, WITH
CC       PROBLABLY FUC AS REDUCING TERMINAL SUGAR.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ---------------------------------------------------------------------
DR   EMBL; X91895; CAA62997.1; -
DR   EMBL; S80643; AAB35894.2; -
DR   EMBL; Y11719; CAA72406.1; -
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain;
KW   Alternative splicing.
FT   SIGNAL        1     18
FT   CHAIN        19    427       LACTADHERIN.
FT   DOMAIN       20     59       EGF-LIKE 1
FT   DOMAIN       62    106       EGF-LIKE 2
FT   DOMAIN      109    265       F5/8 TYPE C 1.
FT   DOMAIN      270    427       F5/8 TYPE C 2.
FT   SITE         85     87       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     24     35       BY SIMILARITY.
FT   DISULFID     29     47       BY SIMILARITY.
FT   DISULFID     49     58       BY SIMILARITY.
FT   DISULFID     66     77       BY SIMILARITY
FT   DISULFID     71     94       BY SIMILARITY
FT   DISULFID     96    105       BY SIMILARITY.
FT   DISULFID    109    265
FT   DISULFID    252    256
FT   DISULFID    270    427
FT   CARBOHYD     27     27       O-LINKED (FUC. . .) (IN PAS-6).
FT   CARBOHYD     34     34       O-LINKED (FUC      ) (IN PAS-7)
FT   CARBOHYD     59     59       N-LINKED (GLCNAC.    .) (HYBRID) (IN PAS-6
FT                                AND PAS-7).
FT   CARBOHYD    227    227       N-LINKED (GLCNAC      ) (HIGH MANNOSE) (IN
FT                                PAS-6).
FT   VARSPLIC    169    221       MISSING (IN SHORT ISOFORM).
FT   CONFLICT     19     19       A -> F (IN REF. 1).
FT   CONFLICT     28     28       L -> Q (IN REF. 1)
SQ   SEQUENCE   427 AA;  47411 MW;  4CBBEE3A1BC4EB24 CRC64;
```

```
  Query Match             78.3%;  Score 36;  DB 1;  Length 427;
  Best Local Similarity   75.0%;  Pred. No. 4.7;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           |||:|||:|
Db     234 NLFDTPLE 241
```

```
RESULT   5
MFGM_PIG
ID   MFGM_PIG        STANDARD;      PRT:   409 AA.
AC   P79385;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Lactadherin (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM) (Sperm
DE   surface protein SP47) (PP47).
GN   MFGE8.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC       SPERMATOZOAN.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y11683; CAA72379.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Glycoprotein; Repeat; EGF-like domain.
FT   DOMAIN        2     41       EGF-LIKE 1.
FT   DOMAIN       44     88       EGF-LIKE 2.
FT   DOMAIN       91    247       F5/8 TYPE C 1.
FT   DOMAIN      252    409       F5/8 TYPE C 2.
FT   SITE         67     69       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID      6     17       BY SIMILARITY.
FT   DISULFID     11     29       BY SIMILARITY.
FT   DISULFID     31     40       BY SIMILARITY.
FT   DISULFID     91    247       BY SIMILARITY.
FT   DISULFID    234    238       BY SIMILARITY.
FT   DISULFID    252    409       BY SIMILARITY.
FT   CARBOHYD     41     41       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC    .) (POTENTIAL).
SQ   SEQUENCE   409 AA;  45725 MW;  B9C07AFB9D2992/A CRC64;
```

```
  Query Match             73.9%;  Score 34;  DB 1;  Length 409;
  Best Local Similarity   75.0%;  Pred. No. 12;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           |||| |:|
Db     216 NLFEVPLE 223
```

```
RESULT   6
MUTA_CAEEL
ID   MUTA_CAEEL     STANDARD;      PRT;   744 AA.
AC   Q23381;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable methylmalonyl-CoA mutase, mitochondrial precursor
DE   (EC 5.4.99.2) (MCM).
GN   ZK1058.1
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Mortimore B.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED, IN MAN, IN THE DEGRADATION OF SEVERAL AMINO
CC       ACIDS, ODD-CHAIN FATTY ACIDS AND CHOLESTEROL VIA PROPIONYL-COA TO
CC       THE TRICARBOXYLIC ACID CYCLE. MCM HAS DIFFERENT FUNCTIONS IN
CC       OTHER SPECIES (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: (R) 2-methyl-3-oxopropanoyl-CoA = succinyl-
CC       CoA.
CC   -!- COFACTOR: ADENOSYLCOBALAMIN (BY SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE METHYLMALONYL-COA MUTASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z35604; CAA84676.1; -.
DR   HSSP; P11653; 1PEQ.
DR   WormPep; ZK1058.1; CE01101.
DR   InterPro; IPR003312; B12-binding.
DR   InterPro; IPR001629; MM_CoA_mutase.
DR   Pfam; PF02310; B12_binding; 1.
DR   Pfam; PF01642; MM_CoA_mutase; 1.
DR   PROSITE; PS00544; METMALONYL_COA_MUTASE; 1.
KW   Mitochondrion; Transit peptide; Isomerase; Vitamin B12; Cobalt.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?    744       PROBABLE METHYLMALONYL-COA MUTASE.
FT   DOMAIN      616    694       COBALAMIN-BINDING (POTENTIAL).
FT   METAL       618    618       COBALT (POTENTIAL).
SQ   SEQUENCE   744 AA;  81669 MW;  4871C1EA93CDF206 CRC64;


  Query Match             73.9%;  Score 34;  DB 1;  Length 744;
  Best Local Similarity   75.0%;  Pred. No. 22;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 LFETPVEA 9
           ||:||:||
Db     641 LFQTPLEA 648


RESULT   7
BLAC_RHOCA
ID   BLAC_RHOCA     STANDARD;      PRT;   293 AA.
AC   P14171;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Beta-lactamase precursor (EC 3.5.2.6) (Penicillinase).
OS   Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1061;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SP108;
RX   MEDLINE=89350885; PubMed=2788410;
RA   Campbell J.I.A., Scaife S., Gibson T., Ambler R.P.;
RT   "The phototrophic bacterium Rhodopseudomonas capsulata sp108 encodes
RT   an indigenous class A beta-lactamase.";
RL   Biochem. J. 260:803-812(1989).
CC   -!- FUNCTION: HYDROLYSES BETA-LACTAMS ANTIBIOTICS. RATES OF
CC       HYDROLYSIS RELATIVE TO BENZYLPENICILLIN (100): AMPICILLIN = 27,
CC       CARBENICILLIN = 25, CLOXACILLIN = 0, CEPHALORIDINE = 4.
CC   -!- CATALYTIC ACTIVITY: A beta lactam + H(2)O = a substituted beta-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Periplasmic.
CC   -!- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15791; CAA33795.1; -.
DR   PIR; S04649; S04649.
DR   HSSP; P00810; 1TEM.
DR   InterPro; IPR001466; Beta_lactam.
DR   InterPro; IPR000871; Beta_lactam_A.
DR   Pfam; PF00144; beta-lactamase; 1.
DR   PRINTS; PR00118; BLACTAMASEA.
DR   PROSITE; PS00146; BETA_LACTAMASE_A; 1.
KW   Hydrolase; Antibiotic resistance; Periplasmic; Signal.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    293       BETA-LACTAMASE.
FT   ACT_SITE     74     74       BY SIMILARITY.
FT   BINDING     238    240       SUBSTRATE (BY SIMILARITY).
SQ   SEQUENCE   293 AA;  31295 MW;  [illegible] CRC64;


  Query Match             73.9%;  Score 34;  DB 1;  Length 293;
  Best Local Similarity   77.8%;  Pred. No. 14;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
             :| ||||||
Db      25 SLAETPVEA 33


RESULT   8
YSV5_CAEEL
ID   YSV5_CAEEL     STANDARD;      PRT;   464 AA.
AC   Q10011;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 51.3 kDa protein T19C3.5 in chromosome III precursor.
GN   T19C3.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Fulton L.;
RL   Submitted (JUN 1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: SOME, TO THE BPI/CFIP/LBP/PLTP FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U28412; AAC46596.1; -.
DR   WormPep; T19C3.5; CE02056.
DR   InterPro; IPR001124; LBP_BPI_CETP.
DR   Pfam; PF01273; LBP_BPI_CETP; 1.
DR   Pfam; PF02886; LBP_BPI_CETP_C; 1.
DR   SMART; SM00328; BPI1; 1.
DR   SMART; SM00329; BPI2; 1.
DR   PROSITE; PS00400; LBP_BPI_CETP; FALSE_NEG.
KW   Hypothetical protein; Signal.
FT   SIGNAL        1     17       POTENTIAL
FT   CHAIN        18    464       HYPOTHETICAL PROTEIN T19C3.5.
SQ   SEQUENCE   464 AA;  51939 MW;  3029CFD72F65E67F CRC64;


  Query Match             71.7%;  Score 33;  DB 1;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NLFETP 6
          ||||||
Db    134 NLFETP 139


RESULT   9
SPKC_SYNY3
ID   SPKC_SYNY3      STANDARD;      PRT;   535 AA.
AC   P74745;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable serine/threonine-protein kinase C (EC 2.7.1.-).
GN   SPKC OR SLR0599.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kamei A., Ikeuchi M.;
RT   "A novel gene, spkC, encodes active Ser/Thr protein kinase in the
RT   motile cyanobacterium Synechocystis sp. PCC 6803.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AB046599; BAB17035.1; -.
DR   EMBL; D90917; BAA18865.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Transferase; Serine/threonine protein kinase; ATP-binding;
KW   Complete proteome.
FT   DOMAIN       12    277       PROTEIN KINASE.
FT   NP_BIND      18     26       ATP (BY SIMILARITY).
FT   BINDING      43     43       ATP (BY SIMILARITY).
FT   ACT_SITE    142    142       BY SIMILARITY.
SQ   SEQUENCE   535 AA;  58141 MW;  533CD9FC0E3D23D8 CRC64;


  Query Match             71.7%;  Score 33;  DB 1;  Length 535;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NLFETP 6
          ||||||
Db    396 NLFETP 401


RESULT  10
NU5M_GADMO
ID   NU5M_GADMO      STANDARD;      PRT;   612 AA.
AC   P55782;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3).
GN   MTND5 OR ND5.
OS   Gadus morhua (Atlantic cod).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidei; Gadidae;
OC   Gadus.
OX   NCBI_TaxID=8049;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NORWEGIAN COASTAL 1;
RX   MEDLINE=96414925; PubMed=8817926;
RA   Johansen S., Bakke I.;
RT   "The complete mitochondrial DNA sequence of Atlantic cod (Gadus
RT   morhua). Relevance to taxonomic studies among codfishes.";
RL   Mol. Mar. Biol. Biotechnol. 5:203-214(1996).
CC   -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X99772; CAA68115.1; -.
DR   InterPro; IPR003918; NADHub_oxdrdctse5.
DR   InterPro; IPR001750; Oxidored_q1.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   Pfam; PF00361; Oxidored_q1; 1.
DR   Pfam; PF00662; Oxidored_q1_N; 1.
DR   PRINTS; PR01434; NADHDHGNASE5.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ   SEQUENCE   612 AA;  68049 MW;  4A639F2F51FFA158 CRC64;


  Query Match             71.7%;  Score 33;  DB 1;  Length 612;
  Best Local Similarity   85.7%;  Pred. No. 30;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NLFETPV 7
          |||.|||
```

```
Db    479 NLFKTPV 485


RESULT  11
Y32K_SSV1
ID   Y32K_SSV1      STANDARD;      PRT:   291 AA.
AC   P20197;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   Hypothetical 33.7 kDa protein (ORF A 291)
OS   Sulfolobus virus-like particle SSV1
OC   Viruses; dsDNA viruses, no RNA stage; Fuselloviridae; Fusellovirus.
OX   NCBI_TaxID=10476;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92024080; PubMed=1426776;
RA   Palm P., Schleper C., Grampp B., Yeats S., McWilliam P., Reiter W.-D.,
RA   Zillig W.;
RT   "Complete nucleotide sequence of the virus SSV1 of the
RT   archaebacterium Sulfolobus shibatae.";
RL   Virology 185:242-250(1991).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X07234; CAA30198.1; -.
DR   PIR; S03230; S03230.
KW   Hypothetical protein.
SQ   SEQUENCE   291 AA;  31669 MW;  84B23AB748DBF2A CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 291;
  Best Local Similarity   55.6%;  Pred. No. 22;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |:| :|:||
Db     101 NIFNSPLEA 109


RESULT  12
GUB2_HORVU
ID   GUB2_HORVU     STANDARD;      PRT:   312 AA.
AC   P12257;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Lichenase II precursor (EC 3.2.1.73) (Endo beta-1,3-1,4 glucanase II)
DE   ((1->3,1->4)-beta-glucanase isoenzyme EII) (Fragment).
OS   Hordeum vulgare (Barley).
OC   Eukaryota, Viridiplantae; Streptophyta, Embryophyta, Tracheophyta,
OC   Spermatophyta, Magnoliophyta, Liliopsida, Poales; Poaceae, Pooideae,
OC   Triticeae; Hordeum.
OX   NCBI_TaxID=4513;
RN   [1]
RP   SEQUENCE OF 1-291 FROM N.A., AND SEQUENCE OF 292-312.
RC   STRAIN=CV. HIMALAYA;
RA   Fincher G.B., Lock P.A., Morgan M.M., Lingelbach K.,
RA   Wettenhall R.E.H., Mercer J.F.B., Brandt A., Thomsen K.K.;
RT   "Primary structure of the (1-3,1-4)-beta-D-glucan 4-glucohydrolase
RT   from barley aleurone.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:2081-2085(1986).
RN   [2]
RP   SEQUENCE OF 7-46.
RC   STRAIN=CV. CLIPPER; TISSUE=Seed;
RA   Woodward J.R., Morgan F.J., Fincher G.B.,
RT   "Amino acid sequence homology in two 1,3;1,4 beta-glucan
RT   endohydrolases from germinating barley (Hordeum vulgare).";
RL   FEBS Lett. 138:198-200(1982).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX   MEDLINE 94199828, PubMed 8146192,
RA   Varghese J.N., Garrett T.P.J., Colman P.M., Chen L., Hoej P.B.,
RA   Fincher G.B.;
RT   "Three-dimensional structures of two plant beta-glucan endohydrolases
RT   with distinct substrate specificities.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:2785-2789(1994).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX   MEDLINE 98123117, PubMed 9452456,
RA   Mueller J.J., Thomsen K.K., Heinemann U.;
RT   "Crystal structure of barley 1,3-1,4-beta-glucanase at 2.0-A
RT   resolution and comparison with Bacillus 1,3-1,4 beta-glucanase.";
RL   J. Biol. Chem. 273:3438-3446(1998).
CC   -!- FUNCTION: FUNCTION IN PLANT CELL WALL HYDROLYSIS DURING
CC       MOBILIZATION OF THE ENDOSPERM IN GERMINATING GRAIN OR DURING THE
CC       GROWTH OF VEGETATIVE TISSUES.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glycosidic linkages
CC       in beta-D-glucans containing 1,3- and 1,4 bonds.
CC   -!- PATHWAY: DEPOLYMERIZATION OF BETA-GLUCAN
CC   -!- SIMILARITY: BELONGS TO FAMILY 17 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M13237; AAA32962.1; -.
DR   PIR; B18891; B18891.
DR   PIR; A25455; A25455.
DR   PDB; 1GHR; 01-NOV-94
DR   PDB; 1AQ0; 11-FEB-98.
DR   InterPro; IPR000490; Glyco_hydro_17.
DR   Pfam; PF00332; Glyco_hydro_17; 1.
DR   PROSITE; PS00587; GLYCOSYL_HYDROL_F17; 1.
KW   Hydrolase; Glycosidase; Signal; Glycoprotein; 3D-structure.
FT   NON_TER       1      1
FT   SIGNAL       <1      6
FT   CHAIN         7    312       LICHENASE II.
FT   ACT_SITE    238    238       NUCLEOPHILE.
FT   ACT_SITE    294    294       PROTON DONOR.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC...).
SQ   SEQUENCE   312 AA;  32774 MW;  7D18C2A6DF522F17 CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 312;
  Best Local Similarity   66.7%;  Pred. No. 24;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |||.| |:|
Db     211 NLFDTIVDA 219


RESULT  13
YOPM_YERPE
ID   YOPM_YERPE     STANDARD;      PRT:   409 AA.
AC   P17778; O68701;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01 MAR 2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Outer membrane protein yopM.
GN   YOPM OR YOP48 OR YPCD1.26C OR Y0059.
OS   Yersinia pestis.
OG   Plasmid pCD1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
```

```
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5;
RX   MEDLINE=89359090; PubMed=2670888;
RA   Leung K.Y., Straley S.C.;
RT   "The yopM gene of Yersinia pestis encodes a released protein having
RT   homology with the human platelet surface protein GPIb alpha ";
RL   J. Bacteriol. 171:4623-4632(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5;
RX   MEDLINE=98427122; PubMed=9744557;
RA   Perry R.D., Straley S.C., Fetherston J.D., Rose D.J., Gregor J.,
RA   Blattner F.R.;
RT   "DNA sequencing and analysis of the low Ca2+ response plasmid pCD1 of
RT   Yersinia pestis KIM5.";
RL   Infect. Immun. 66:4611-4623(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5;
RX   MEDLINE=98422454; PubMed=9748454;
RA   Hu P., Elliott J., McCready P., Skowronski E., Garnes J.,
RA   Kobayashi A., Brubaker R.R., Garcia E.;
RT   "Structural organization of virulence-associated plasmids of Yersinia
RT   pestis.";
RL   J. Bacteriol. 180:5192-5202(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE AND SECRETED.
CC   -!- SIMILARITY: CONTAINS 12 LEUCINE-RICH REPEATS (LRR).
CC   -!- SIMILARITY: TO S.FLEXNERI IPAH 7.8
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ---------------------------------------------------------------------------
DR   EMBL; M25819; AAA27670.1; -.
DR   EMBL; AF074612; AAC69806.1; -.
DR   EMBL; AF053946; AAC62580.1; -.
DR   EMBL; AL117189; CAB54903.1; -.
DR   PIR; A33950; A33950.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003591; LRR_out
DR   Pfam; PF00560; LRR; 6.
DR   SMART; SM00370; LRR; 12.
KW   Outer membrane; Repeat; Leucine-rich repeat; Antigen; Plasmid;
KW   Complete proteome.
FT   REPEAT       69     90       LRR 1.
FT   REPEAT       91    110       LRR 2.
FT   REPEAT      111    129       LRR 3.
FT   REPEAT      130    152       LRR 4.
FT   REPEAT      153    177       LRR 5.
FT   REPEAT      179    194       LRR 6.
FT   REPEAT      195    217       LRR 7.
FT   REPEAT      236    256       LRR 8.
FT   REPEAT      257    281       LRR 9.
FT   REPEAT      296    316       LRR 10.
FT   REPEAT      317    341       LRR 11.
FT   REPEAT      356    381       LRR 12.
FT   CONFLICT    142    183       MISSING (IN REF. 1).
SQ   SEQUENCE   409 AA;  46293 MW;  E224090b7fA54951 CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 409;
  Best Local Similarity   77.8%;  Pred. No. 32,
  Matches    7;  Conservative    0,  Mismatches    2,  Indels    0,  Gaps    0;

Qy       1 NLFETPVEA 9
           || | ||||
Db      22 NLTEMPVEA 30


RESULT  14
NU85_YEAST
ID   NU85_YEAST     STANDARD;      PRT;   744 AA.
AC   P46673;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Nucleoporin NUP85 (Nuclear pore protein NUP85)
GN   NUP85 OR RAI9 OR YGR042W OR J1624.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FY23;
RX   MEDLINE=96413892; PubMed=8811938;
RA   Goldstein A.L., Snay C.A., Heath C.V., Cole C.N.,
RT   "Pleiotropic nuclear defects associated with a conditional allele of
RT   the novel nucleoporin Rat9p/Nup85p.";
RL   Mol. Biol. Cell 7:917-934(1996).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 31-39; 123-138 AND 708-718.
RX   MEDLINE=96152656; PubMed=8565072;
RA   Siniossoglou S., Wimmer C., Rieger M., Doye V., Tekotte H., Weise C.,
RA   Emig S., Segref A., Hurt E.C.;
RT   "A novel complex of nucleoporins, which includes Sec13p and a Sec13p
RT   homolog, is essential for normal nuclear pores.";
RL   Cell 84:265-275(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=95397545; PubMed=7668047;
RA   Huang M.-E., Chuat J.-C., Galibert F.;
RT   "Analysis of a 42.5 kb DNA sequence of chromosome X reveals three
RT   tRNA genes and 14 new open reading frames including a gene most
RT   probably belonging to the family of ubiquitin-protein ligases.";
RL   Yeast 11:775-781(1995).
CC   -!- FUNCTION: REQUIRED FOR EFFICIENT MRNA EXPORT FROM THE NUCLEUS TO
CC       THE CYTOPLASM AND FOR CORRECT NUCLEAR PORE BIOGENESIS.
CC   -!- SUBUNIT: INTERACTS WITH NUP84, NUP120, SEC13 AND SEH1.
CC   -!- SUBCELLULAR LOCATION: Nuclear pore complex.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U36469; AAB48943.1; -.
DR   EMBL; X90995; CAA62481.1; -.
DR   EMBL; Z49542; CAA89569.1; -.
DR   EMBL; U36344; AAA88744.1; -.
DR   SGD; S0003803; NUP85.
KW   Nuclear protein; Transport.
FT   CONFLICT    101    101       S -> T (IN REF. 1).
```

```
FT   CONFLICT    471    471       P -> L (IN REF. 1).
SQ   SEQUENCE   744 AA;  84897 MW;  4C0A4AD3DB57A023 CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 744;
  Best Local Similarity   66.7%;  Pred. No. 59;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           ||:|| :||
Db     708 NLYETIIEA 716


RESULT  15
VGR3_MOUSE
ID   VGR3_MOUSE     STANDARD;      PRT;  1363 AA.
AC   P35917;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Vascular endothelial growth factor receptor 3 precursor (EC 2.7.1.112)
DE   (VEGFR-3) (Tyrosine-protein kinase receptor FLT4).
GN   FLT4 OR FLT 4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=93330572; PubMed=8393164;
RA   Finnerty H., Kelleher K., Morris C.F., Bean K., Merberg D.M.,
RA   Kriz R., Morris J.C., Sookdeo H., Turner K.J., Wood C.R.;
RT   "Molecular cloning of murine FIT and FLT4.";
RL   Oncogene 8:2293-2298(1993).
CC   -!- FUNCTION: RECEPTOR FOR VEGF-C. HAS A TYROSINE-PROTEIN KINASE
CC       ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN ADULT LUNG AND LIVER, AND IN
CC       FETAL LIVER, BRAIN, INTESTINE AND PLACENTA.
CC   -!- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-
CC       PROTEIN KINASES.
CC   -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2 TYPE DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ----------------------------------------------------------------------
DR   EMBL; L07296; AAA40077.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:95561; Flt4.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR001824; Receptor_tyr_kin_III.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00047; ig; 6.
DR   Pfam; PF00069; pkinase; 2.
DR   SMART; SM00410; IG_like; 3.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine protein kinase; Phosphorylation; ATP-binding;
KW   Receptor; Transmembrane; Signal; Immunoglobulin domain; Repeat;
KW   Glycoprotein.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   1363       VASCULAR ENDOTHELIAL GROWTH FACTOR
FT                                RECEPTOR 3.
FT   DOMAIN       25    775       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    776    797       POTENTIAL.
FT   DOMAIN      798   1363       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       44    118       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      151    213       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      245    317       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      351    403       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      438    541       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      571    660       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      692    758       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN      845   1173       PROTEIN KINASE.
FT   NP_BIND     851    859       ATP (BY SIMILARITY).
FT   BINDING     879    879       ATP (BY SIMILARITY).
FT   ACT_SITE   1037   1037       BY SIMILARITY.
FT   DISULFID     51    111       POTENTIAL.
FT   DISULFID    158    206       POTENTIAL.
FT   DISULFID    252    310       POTENTIAL.
FT   DISULFID    445    534       POTENTIAL.
FT   DISULFID    578    653       POTENTIAL.
FT   DISULFID    699    751       POTENTIAL.
FT   MOD_RES    1068   1068       PHOSPHORYLATION (AUTO?) (BY SIMILARITY).
FT   CARBOHYD     33     33       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    166    166       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    251    251       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    411    411       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    527    527       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    582    582       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    594    594       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    683    683       N-LINKED (GLCNAC      ) (POTENTIAL).
FT   CARBOHYD    690    690       N-LINKED (GLCNAC      ) (POTENTIAL).
FT   CARBOHYD    758    758       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1363 AA;  153015 MW;  F1BF8A2BDEF99BE9 CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 1363;
  Best Local Similarity   66.7%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           :|| ||:||
Db     614 HLFATPLEA 622



Search completed: September  5, 2002, 15:31:42
Job time: 477 sec
```

Thu Ser

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:41:03 ; Search time 122.88 Seconds
(without alignments)
12.673 Million cell updates/sec

Title: US-09-744-804-38
Perfect score: 46
Sequence: 1 NLFETPVEA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172991929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified *
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	46	100.0	335	4	Q9BTL9	Q9btl9 homo sapien
2	38	82.6	2261	4	Q96T85	Q96t85 homo sapien
3	38	82.6	2261	4	Q96S56	Q96s56 homo sapien
4	37	80.4	393	16	Q982T3	Q982t3 rhizobium l
5	36	78.3	78	4	P78328	P78328 homo sapien
6	35	76.1	392	8	Q9UV31	Q9uv31 emericella
7	35	76.1	566	3	Q10295	Q10295 schizosacch
8	35	76.1	658	3	Q74798	Q74798 schizosacch
9	34	73.9	79	2	Q9ZHV4	Q9zhv4 helicobacte
10	34	73.9	177	10	Q9SVB2	Q9svb2 arabidopsis
11	34	73.9	261	17	O30097	O30097 archaeoglob
12	34	73.9	439	10	Q941M0	Q941m0 brassica ol
13	34	73.9	714	10	Q9AV92	Q9av92 oryza sativ
14	34	73.9	762	3	Q9UTE0	Q9ute0 schizosacch
15	33	71.7	54	16	Q983T3	Q983t3 rhizobium l
16	33	71.7	265	2	Q9ZP63	Q9zp63 campylobact
17	33	71.7	269	16	Q9PMS4	Q9pms4 campylobact
18	33	71.7	343	5	Q9XWP5	Q9xwp5 caenorhabdi
19	33	71.7	405	17	Q9HS37	Q9hs37 halobacteri
20	33	71.7	714	16	Q9A5S3	Q9a5s3 caulobacter
21	33	71.7	1360	3	Q9P4Z2	Q9p4z2 neurospora
22	32	69.6	179	16	Q9A8I6	Q9a8i6 caulobacter
23	32	69.6	196	16	Q9K7Q5	Q9k7q5 bacillus ha
24	32	69.6	334	10	Q02345	Q02345 hordeum vul
25	32	69.6	334	10	Q42518	Q42518 avena sativ
26	32	69.6	334	10	Q40686	Q40686 oryza sativ
27	32	69.6	334	10	Q07556	Q07556 triticum ae
28	32	69.6	376	10	Q9LX81	Q9lx81 arabidopsis
29	32	69.6	383	10	Q9CAP3	Q9cap3 arabidopsis
30	32	69.6	418	3	Q9USM3	Q9usm3 schizosacch
31	32	69.6	422	17	Q9HPT0	Q9hpt0 halobacteri
32	32	69.6	440	10	Q941J5	Q941j5 arabidopsis
33	32	69.6	571	10	Q04250	Q04250 arabidopsis
34	32	69.6	707	10	Q9FHM4	Q9fhm4 arabidopsis
35	32	69.6	772	5	Q9CK73	Q9ck73 leishmania
36	32	69.6	783	1	Q50247	Q50247 methanosare
37	32	69.6	790	5	Q20949	Q20949 caenorhabdi
38	32	69.6	820	12	Q69375	Q69375 mouse cytom
39	32	69.6	1092	11	Q91ZT0	Q91zt0 rattus norv
40	32	69.6	1363	11	Q91ZT1	Q91zt1 rattus norv
41	32	69.6	1673	1	Q977V5	Q977v5 methanosare
42	32	69.6	1960	5	Q9NGU2	Q9ngu2 leishmania
43	32	69.6	1960	5	Q9NE08	Q9ne08 leishmania
44	31	67.4	55	8	Q9T2D7	Q9t2d7 bos taurus
45	31	67.4	79	2	Q9ZHU2	Q9zhu2 helicobacte

ALIGNMENTS

RESULT 1
Q9BTL9
ID Q9BTL9 PRELIMINARY; PRT; 335 AA.
AC Q9BTL9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO MILK FAT GLOBULE-EGF FACTOR 8 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia, Eutheria, Primates, Catarrhini, Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA.;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
DR EMBL; BC003610; AAH03610.1; -.
DR HSSP; P08709; 1BF9.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000421; FA58_C.
DR PRINTS; PR00010; EGFBLOOD.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01285; FA58C_1; 1.
DR PROSITE; PS01286; FA58C_2; 2.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 335 AA; 37523 MW; 22E84E55B78E9E2F CRC64;

Query Match 100.0%; Score 46; DB 4; Length 335;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NLFETPVEA 9
           |||||||||
Db     194 NLFETPVEA 202


RESULT   2
Q96T85
ID   Q96T85      PRELIMINARY;      PRT;  2261 AA.
AC   Q96T85;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ATP-BINDING CASSETTE 1 SUB-FAMILY A MEMBER 1.
GN   ABCA1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21251004; PubMed=11352567;
RA   Qin Y., Cavelier L., Chiu S., Yang X., Rubin E., Cheng J.-F.,
RT   "Human and Mouse ABCA1 Comparative Sequencing and Transgenesis Studies
RT   Revealing Novel Regulatory Sequences.";
RL   Genomics 73:66-76(2001).
DR   EMBL; AF287262; AAK43526.1; -.
KW   ATP binding.
SQ   SEQUENCE   2261 AA;  254238 MW;  248DF04C8FF041A5 CRC64;


  Query Match             82.6%;  Score 38;  DB 4;  Length 2261;
  Best Local Similarity   87.5%;  Pred. No. 48;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           ||||:|||
Db     808 NLFESPVE 815


RESULT   3
Q96S56
ID   Q96S56      PRELIMINARY;      PRT;  2261 AA.
AC   Q96S56;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ABCA1.
GN   ABCA1.
OS   Homo sapiens (Human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tanaka A.R., Abe-Dohmae S., Arakawa R., Sadanami K., Kidera A.,
RA   Kioka N., Amachi T., Yokoyama S., Ueda K.;
RT   "A new topological model of functional human ABCA1-Signal peptide
RT   cleavage and glycosylation of a large extracellular domain.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB055982; BAB63210.1; -.
SQ   SEQUENCE   2261 AA;  254208 MW;  21A3C8B24F51D0D CRC64;


  Query Match             82.6%;  Score 38;  DB 4;  Length 2261;
  Best Local Similarity   87.5%;  Pred. No. 48;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           ||||:|||
Db     808 NLFESPVE 815


RESULT   4
```

```
Q982T3
ID   Q982T3      PRELIMINARY;      PRT;   393 AA.
AC   Q982T3;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   MLR8507 PROTEIN.
GN   MLR8507.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; Alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP003014; BAB54373.1; -.
DR   InterPro; IPR002052; N6_Mtase.
DR   PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   393 AA;  43166 MW;  427F5CC3BBF59B6B CRC64;


  Query Match             80.4%;  Score 37;  DB 16;  Length 393;
  Best Local Similarity   77.8%;  Pred. No. 12;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           ||:||| ||
Db     173 NLYETPPEA 181


RESULT   5
P78328
ID   P78328      PRELIMINARY;      PRT;    78 AA.
AC   P78328;
DT   01 MAY 1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BP47 PROTEIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Ensslin M.;
RL   Thesis (1996), University of Hannover, Hannover, FRG.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=96200817; PubMed=8641340
RA   Ensslin M., Vogel T., Calvete J.J., Thole H.H., Schmidtke J.,
RA   Matsuda T., Toepfer-Petersen E.;
RT   "Molecular cloning and characterization of BP47, a novel boar sperm-
RT   associated zona pellucida-binding protein homologous to a family of
RT   mammalian secretory proteins.";
RL   Biol. Reprod. 54:1057-1064(1996).
DR   EMBL; Y11718; CAA72405.1; -.
DR   InterPro; IPR000421; FA58_C.
DR   PROSITE; PS01286; FA58C_2; 1.
FT   NON_TER       1      1
FT   NON_TER      78     78
SQ   SEQUENCE   78 AA;  8696 MW;  7EBBDB4B594B5DDA CRC64;
```

Query Match 78.3%; Score 36; DB 4; Length 78;
Best Local Similarity 75.0%; Pred. No. 3.3;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLFETPVE 8
||||:||:|
Db 28 NLFDTPLE 35

RESULT 6
Q9UV31
ID Q9UV31 PRELIMINARY; PRT; 392 AA.
AC Q9UV31;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 60S RIBOSOMAL PROTEIN L3
GN RPL3.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC4;
RX MEDLINE=21100459; PubMed=11179686;
RA Jeong H., Cho G., Han K., Kim J., Min Han D., Jahng K., Chae K.;
RT "Differential expression of house-keeping genes of Aspergillus
RT nidulans during sexual development.";
RL Gene 265:215-219(2001).
DR EMBL; AF198447; AAF15609.1; -.
DR InterPro; IPR000597; Ribosomal_L3.
DR Pfam; PF00297; Ribosomal_L3; 1.
DR ProDom; PD001374; Ribosomal_L3; 1.
DR PROSITE; PS00474; RIBOSOMAL_L3; UNKNOWN_1.
KW Ribosomal protein.
SQ SEQUENCE 392 AA; 44408 MW; 2C20B48CCCFBA4BE CRC64;

Query Match 76.1%; Score 35; DB 3; Length 392;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLFETPVE 8
|||| |:|
Db 198 NLFEKPIE 205

RESULT 7
Q10295
ID Q10295 PRELIMINARY; PRT; 566 AA.
AC Q10295;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POLYMERASE.
GN PLA1.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96292264; PubMed=8692700;
RA Ohnacker M., Miniviella Sebastia L., Keller W.,
RT "The Schizosaccharomyces pombe pla1 gene encodes a poly(A) polymerase
RT and can functionally replace its Saccharomyces cerevisiae homologue.";
RL Nucleic Acids Res. 24:2585-2591(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; X79705; CAA56141.1; -.
DR EMBL; AL031678; CAA21808.1; -.
DR HSSP; P25500; 1F5A.
DR InterPro; IPR002934; NTP_transf.
DR InterPro; IPR001201; PAP_25A_core.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF01909; NTP_transf_2; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
SQ SEQUENCE 566 AA; 64108 MW; E8E0B41B5AAFD3D5 CRC64;

Query Match 76.1%; Score 35; DB 3; Length 566;
Best Local Similarity 66.7%; Pred. No. 47;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLFETPVEA 9
||||:|.|:
Db 36 NLFESPAES 44

RESULT 8
O74798
ID O74798 PRELIMINARY; PRT; 658 AA.
AC O74798;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 57.9 KDA PROTEIN C2D10.04 IN CHROMOSOME II.
GN SPBC2D10.04.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Taylor K., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: TO YEAST YJL084L.
DR EMBL; AL031788; CAA21162.1; -.
SQ SEQUENCE 658 AA; 72731 MW; 40EB0B6BB970B4F7 CRC64;

Query Match 76.1%; Score 35; DB 3; Length 658;
Best Local Similarity 87.5%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLFETPVE 8
|||||| |
Db 423 NLFETPDE 430

RESULT 9
Q9ZHV4
ID Q9ZHV4 PRELIMINARY; PRT; 79 AA.
AC Q9ZHV4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE FEPC (FRAGMENT).
GN FEPC.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F20;
RX MEDLINE=98454456; PubMed=9780260;

```
RA   Ito Y., Aruma T., Ito S., Sato H., Miyaji H., Yamazaki Y., Kohli Y.,
RA   Kuriyama M.;
RT   "Full-length sequence analysis of the vacA gene from cytotoxic and
RT   noncytotoxic Helicobacter pylori";
RL   J. Infect. Dis. 178:1391-1398(1998).
DR   EMBL; AF049619; AAD04260.1.
FT   NON_TER       1       1
SQ   SEQUENCE   79 AA;  9102 MW;  5337EB2167EA9A0F CRC64;


  Query Match             73.9%;  Score 34;  DB 2;  Length 79;
  Best Local Similarity   75.0%;  Pred. No. 9.1;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LFETPVEA 9
           |:|||:||
Db      60 LYETPLEA 67


RESULT  10
Q9SVD2
ID   Q9SVD2      PRELIMINARY;      PRT;   177 AA.
AC   Q9SVD2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 20.3 KDA PROTEIN
GN   F2206_110.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Purnelle B., Boutry M., Goffeau A., Mewes H.W., Rudd S., Lemcke K.,
RA   Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project.
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL050300; CAB43421.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   177 AA;  20208 MW;  B4BFD8B94ADC8E9C CRC64;


  Query Match             73.9%;  Score 34;  DB 10;  Length 177;
  Best Local Similarity   77.8%;  Pred. No. 22;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           || | ||||
Db      66 NLIELPVEA 74


RESULT  11
O30097
ID   O30097      PRELIMINARY;      PRT;   261 AA.
AC   O30097;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   UBIQUINONE/MENAQUINONE BIOSYNTHESIS METHYLTRANSFERASE (UBIE).
GN   AF0140.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
DR   EMBL; AE001096; AAB91083.1; -.
DR   TIGR; AF0140; -.
DR   InterPro; IPR001601; Meth-transf.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR004033; ubiE_COQ5_methyltransf.
DR   Pfam; PF01209; Ubie_methyltran; 1.
KW   Hypothetical protein; Transferase; Methyltransferase; Ubiquinone;
KW   Complete proteome.
SQ   SEQUENCE   261 AA;  30256 MW;  E1FBA6AE22089F5B CRC64;


  Query Match             73.9%;  Score 34;  DB 17;  Length 261;
  Best Local Similarity   75.0%;  Pred. No. 33;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVE 8
           |||||:|
Db      35 DLFETPME 42


RESULT  12
Q941M0
ID   Q941M0      PRELIMINARY;      PRT;   439 AA.
AC   Q941M0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   2-OXOGLUTARATE-DEPENDENT DIOXYGENASE
OS   Brassica oleracea (Cauliflower).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3712;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Li G., Quiros C.F.;
RT   "Cloning of two major genes involved in aliphatic glucosinolate
RT   synthesis in Brassica oleracea.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY044425; AAK95851.1; -.
KW   Dioxygenase.
SQ   SEQUENCE   439 AA;  48313 MW;  B7DB2014AE923F3E CRC64;


  Query Match             73.9%;  Score 34;  DB 10;  Length 439;
  Best Local Similarity   87.5%;  Pred. No. 59;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LFETPVEA 9
           ||| ||||
Db      64 LFELPVEA 71


RESULT  13
Q9AV92
ID   Q9AV92      PRELIMINARY;      PRT;   714 AA.
AC   Q9AV92;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
```

```
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   KINASE-LIKE PROTEIN.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. AKITAKOMACHI; TISSUE=LEAF;
RA   Sawada K., Iwata M.;
RT   "Oryza sativa kinase like protein.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB060276; BAB41205.1; -.
DR   InterPro; IPR001487; Bromodomain.
DR   Pfam; PF00439; bromodomain; 1.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
SQ   SEQUENCE   714 AA;  78710 MW;  4F2287942E1C99E9 CRC64;

  Query Match             73.9%;  Score 34;  DB 10;  Length 714;
  Best Local Similarity   55.6%;  Pred. No. 1e+02;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |:|::|||:|
Db     185 NIFDSPVDA 193

RESULT 14
Q9UTE0
ID   Q9UTE0      PRELIMINARY;      PRT;   762 AA.
AC   Q9UTE0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 87.5 KDA PROTEIN C222.14C IN CHROMOSOME I.
GN   SPAC222.14C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Hamlin N., Churcher C.M., McDougall R.C., Rajandream M.A.,
RA   Barrell B.G.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO YEAST YOR165W
DR   EMBL; AL132798; CAB60706.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM    683    703       POTENTIAL.
SQ   SEQUENCE   762 AA;  87510 MW;  3ADDA2FC8F8E813A CRC64;

  Query Match             73.9%;  Score 34;  DB 3;  Length 762;
  Best Local Similarity   66.7%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           |:||| :||
Db     328 NMFETVIEA 336

RESULT 15
Q983T3
ID   Q983T3      PRELIMINARY;      PRT;    54 AA.
AC   Q983T3;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   MSR8189 PROTEIN.
GN   MSR8189.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP003013; BAB53797.1; -.
KW   Complete proteome.
SQ   SEQUENCE   54 AA;  5414 MW;  C807E2EC4758C859 CRC64;

  Query Match             71.7%;  Score 33;  DB 16;  Length 54;
  Best Local Similarity   66.7%;  Pred. No. 9.9;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NLFETPVEA 9
           :||::||||
Db      37 DLFKSPVEA 45

Search completed: September  5, 2002, 15:31:04
Job time: 469 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:26:31 ; Search time 158.43 Seconds
(without alignments)
6.310 Million cell updates/sec

Title: US-09-744-804-39
Perfect score: 54
Sequence: 1 GLQHWVPEL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	54	100.0	9	21	AAY82844	Lactadherin (BA-46
2	54	100.0	387	16	AAR77252	HMFG 46 kDa antige
3	54	100.0	387	21	AAY94453	Human lactadherin
4	40	74.1	217	22	AAG75021	Human colon cancer
5	39	72.2	252	22	AAE01423	Human secreted pro
6	39	72.2	320	20	AAW98891	Peptide SEQ ID NO:
7	39	72.2	426	21	AAY94454	Mouse lactadherin
8	38	70.4	22	22	ABB39426	Peptide #6932 enco
9	38	70.4	22	22	ABB24200	Protein #6199 enco
10	38	70.4	22	22	AAM60107	Human brain expres
11	38	70.4	22	22	AAM72718	Human bone marrow
12	38	70.4	22	22	AAM19701	Peptide #6135 enco
13	38	70.4	22	22	AAM72949	Peptide #6896 enco
14	38	70.4	[illegible]	22	ABB69244	Drosophila melanog
15	38	70.4	4019	22	AAE13839	Human lung tumour-
16	37	68.5	264	19	AAW48615	Escherichia coli m
17	37	68.5	265	17	AAY08234	E. coli RTEM beta-
18	37	68.5	265	20	AAY49892	RTEM with beta-glo
19	37	68.5	265	22	AAE06924	E. coli TEM-1 beta
20	37	68.5	551	22	AAB88436	Human membrane or
21	37	68.5	1016	21	AAB41524	Human ORFX ORF1288
22	37	68.5	3079	15	AAP59926	CAP protein Ira2.
23	36	66.7	15	22	AAP99784	Human SHC protein
24	36	66.7	29	21	AAB51434	Human secreted pro
25	36	66.7	29	21	AAB51442	Human secreted pro
26	36	66.7	53	22	ABB03375	Human musculoskele
27	36	66.7	66	22	AAM80508	Human haematologic
28	36	66.7	66	22	AAM81129	Human haematologic
29	36	66.7	91	22	ABG13244	Novel human diagno
30	36	66.7	155	20	AAY60570	Human normal bladd
31	36	66.7	159	22	AAU40589	Propionibacterium
32	36	66.7	165	22	AAM80655	Human haematologic
33	36	66.7	177	22	AAU39525	Propionibacterium
34	36	66.7	191	22	ABG27674	Novel human diagno
35	36	66.7	194	22	AAM80742	Human haematologic
36	36	66.7	194	22	AAM80755	Human haematologic
37	36	66.7	194	22	AAM81208	Human haematologic
38	36	66.7	196	18	AAW10505	Alphalalpha2 regio
39	36	66.7	196	18	AAW10508	Alphalalpha2 regio
40	36	66.7	231	22	AAB69597	Swine leukocyte an
41	36	66.7	231	22	AAB69598	Swine leukocyte an
42	36	66.7	231	22	AAB69599	Swine leukocyte an
43	36	66.7	232	22	[illegible]	Swine leukocyte an
44	36	66.7	232	22	[illegible]	Swine leukocyte an
45	36	66.7	232	22	[illegible]	Swine leukocyte an

ALIGNMENTS

RESULT 1
AAY82844
ID AAY82844 standard; peptide; 9 AA.
XX
AC AAY82844;
XX
DT 19-JUN-2000 (first entry)
XX
DE Lactadherin (BA-46) peptide fragment (tumour associated antigen).
XX
KW Tumour associated antigen peptide; TAA; cancer; carcinoma;
KW treatment; prevention; cure; anti tumour vaccine; metastases;
KW breast; bladder; prostate; pancreas; ovary; thyroid; colon;
KW stomach; carcinoma; MHC Class I; HLA-A2; human;
KW Major Histocompatibility Complex; uroplakin;
KW prostate specific antigen; prostate specific membrane antigen;
KW prostate acid phosphatase; mucin; lactadherin;
KW teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRIPTO-1.
XX
OS Homo sapiens.
XX
PN WO200006723-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-IL00417.
XX
PR 30-JUL-1998; 98IL-0125608.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (BIOT-) BIO-TECHNOLOGY GEN CORP
XX
PI Eisenbach L, Carmon L, Tirosh B, Bar haim E, Paz A, Fridkin M;
PI Fitzer-attas C;

```
XX
DR   WPI; 2000-205463/18.
XX
PT   Tumor associated antigen peptides, especially derived from uroplakin,
PT   useful as vaccines to prevent or cure cancers including breast,
PT   bladder, prostate, pancreas, ovary, thyroid, colon and stomach -
XX
PS   Claim 17; Page 100; 113pp; English.
XX
CC   Tumour associated antigen peptides (TAA) may be used for the
CC   treatment, prevention and cure of cancer or cancer metastases. The
CC   cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
CC   colon, stomach, head or neck cancer or a carcinoma. The tumour
CC   associated antigens are presentable to the immune system by HLA-A2
CC   molecules and are generally between 8 to 10 amino acids in length.
CC   The amino acids located at positions 2 and 9 of the tumour associated
CC   antigens are the anchor residues which participate in the binding to
CC   MHC class I molecules, more specifically HLA-A2. More tumour
CC   associated antigens are described in GENESEQ records AAY82805-Y82882.
CC   Those tumour associated antigens described in records AAY82806-Y82824
CC   and AAY82855-YY82869 are derived from Uroplakin, such as Uroplakin II,
CC   Uroplakin Ia, Uroplakin III and Uroplakin Ib. Those described in
CC   records AAY82825-Y82829 are derived from prostate specific antigen
CC   (PSA). Those described in records AAY82830 Y82835 are derived from
CC   prostate specific membrane antigen (PSMA). Those described in
CC   records YY82836 AAY82839 are derived from prostate acid phosphotase
CC   (PAP). Those described in records AAY82840 Y82846 are derived
CC   from Lactadherin (BA-46). Those described in records AAY82847-Y82854
CC   are derived from Mucin and those described in records AAY82871-Y82882
CC   are derived from Teratocarcinoma derived growth factor (CRIPTO-1).
XX
SQ   Sequence   9 AA;


  Query Match             100.0%;  Score 54;  DB 21;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           |||||||||
Db       1 glqhwvpel 9


RESULT   2
AAR77252
ID   AAR77252 standard; Protein; 387 AA.
XX
AC   AAR77252;
XX
DT   21-NOV-1995  (first entry)
XX
DE   HMFG 46 kDa antigen.
XX
KW   HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
KW   epithelium; tumor; breast cancer; monoclonal antibody; MAb
XX
OS   Homo sapiens.
XX
PN   WO9515171-A.
XX
PD   08-JUN-1995.
XX
PF   05-DEC-1994;   94WO-US13967.
XX
PR   03-DEC-1993;   93US-0162402.
XX
PA   (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI   Ceriani RL,  Larocca DJ,  Peterson JA;
XX
DR   WPI; 1995-215151/28.
DR   N-PSDB; AAQ91198.
XX
PT   46 kD apparent molecular weight human milk fat globule antigen -
PT   used in assays to determine the presence of a cancerous tumour of
PT   epithelial origin, and in a vaccine against neoplastic tumours
XX
PS   Claim 6; Page 46-47; 68pp; English.
XX
CC   A complete cDNA sequence for the 46 kDa HMFG antigen, a major
CC   component of the apical surface of the normal breast epithelial
CC   cell, was obtained by PCR and RACE methods. cDNA clones can be
CC   used to prepare MAbs for use in immunotherapy, immunohistopathology,
CC   prognosis, diagnosis, imaging and therapy. Recombinant antigen can
CC   be expressed in prokaryotic or (glycosylated) in eukaryotic cells.
XX
SQ   Sequence   387 AA;


  Query Match             100.0%;  Score 54;  DB 16;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.17;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           |||||||||
Db      97 glqhwvpel 105


RESULT   3
AAY94453
ID   AAY94453 standard; Protein; 387 AA.
XX
AC   AAY94453;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Human lactadherin protein.
XX
KW   Human; lactadherin; MGF-E8; anti-tumour; immune response;
KW   exosome; dendritic cell.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Secretion_signal
FT   Protein         24..387
FT                   /label= Lactadherin
FT   Binding-site    46..48
FT                   /label= Integrin_binding_site
XX
PN   EP1004664-A1.
XX
PD   31-MAY-2000.
XX
PF   24-NOV-1998;   98EP-0402925.
XX
PR   24-NOV-1998;   98EP-0402925.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (CURI-) INST CURIE.
XX
DR   WPI; 2000-352597/31.
DR   N-PSDB; AAA27140.
XX
PT   Chimeric isolated (human) lactadherin polypeptide that functions as an
PT   adaptor of cross-priming to eliminate pathogenic antigens, e.g. in
PT   tumors -
XX
PS   Example 3; Page 12; 20pp; English.
XX
CC   Lactadherin protein was found in exosomes produced by dendritic cells.
CC   The protein is involved in the phagocytosis of particulate antigens by
CC   dendritic cells. Exosomes produced by dendritic cells exposed to
CC   tumour antigens induce potent immune responses. Lactadherin or variants
CC   of it may be used in the mediation of an immune response. Variants of
```

```
CC   lactadherin may be used for inhibition and/or stimulation of the
CC   cross-priming of antigens and stimulation of the phagocytosis of
CC   antigens by dendritic cells. Compositions derived from lactadherin can
CC   also be used to monitor an immune response, more specifically a CTL
CC   (cytotoxic T lymphocyte) response and also to produce CTLs specific
CC   for a selected antigen. The present sequence is the human
CC   lactadherin protein.
XX
SQ   Sequence   387 AA;


  Query Match             100.0%;  Score 54;  DB 21;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.17;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           |||||||||
Db      97 glqhwvpel 105


RESULT   4
AAG75021
ID   AAG75021 standard; Protein; 217 AA.
XX
AC   AAG75021;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen protein SEQ ID NO:5785
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235257/24.
DR   N-PSDB; AAH34426.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 11; Page 7304-7305; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive protein or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer associated P,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAP77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7021 and 7022.
XX
SQ   Sequence   217 AA;


  Query Match             74.1%;  Score 40;  DB 22;  Length 217;
  Best Local Similarity   75.0%;  Pred. No. 26;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           ||||:| |
Db     133 lqhwlprl 200


RESULT   5
AAE01423
ID   AAE01423 standard; Protein; 252 AA.
XX
AC   AAE01423;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Human secreted protein variant, SEQ ID NO:147.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW   psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW   inflammation; neurological disorder; Alzheimer's disease; food additive;
KW   angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW   pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW   cell culture; chemotaxis; vulnerary; binding partner identification;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 162
FT                   /note= "Corresponds to any of the naturally occuring
FT                   L-amino acids"
FT   Misc-difference 222
FT                   /note= "Corresponds to any of the naturally occuring
FT                   L-amino acids"
FT   Misc difference 236
FT                   /note= "Corresponds to any of the naturally occuring
FT                   L-amino acids"
FT   Misc difference 240
FT                   /note= "Corresponds to any of the naturally occuring
FT                   L-amino acids"
XX
PN   WO200134629-A1.
XX
PD   17-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-US30654.
XX
PR   12-NOV-1999;   99US-0164835.
PR   27-JUL-2000; 2000US-0221142.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Wei P,  Fikcella M,  Baker KP;
XX
DR   WPI; 2001-387779/32.
XX
PT   New nucleic acid encoding one of 21 human secreted proteins for
PI   diagnosing, preventing, treating or ameliorating medical conditions,
PT   such as autoimmune disease and cancer, and used as a food additive or
PI   preservative -
XX
PS   Disclosure; Page 476-477; 490pp; English.
```

```
XX
CC   AAD05220-AAD05282 represent cDNAs corresponding to 21 human secreted
CC   protein genes, and AAE01352 AAE01413 represent the proteins they encode.
CC   AAE01415-AAE01433 represent human secreted protein fragments or variants.
CC   The secreted proteins and their genes are useful for preventing, treating
CC   or ameliorating medical conditions, e.g., by protein or gene therapy.
CC   Pathological conditions can be diagnosed by determining the amount of the
CC   new protein in a sample or by determining the presence of mutations in
CC   the new genes. Specific uses are described for each of the 21 genes,
CC   based on the tissues in which they are most highly expressed, and include
CC   developing products for the diagnosis or treatment of proliferative
CC   disorders, cancer, tumours, foetal and developmental abnormalities,
CC   haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC   diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC   neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC   cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC   psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC   angiogenic disorders, kidney disorders, gastrointestinal disorders,
CC   pregnancy-related disorders, endocrine disorders, and infections. The
CC   proteins can also be used to aid wound healing and epithelial cell
CC   proliferation, to prevent skin aging due to sunburn, to maintain organs
CC   before transplantation, for supporting cell culture of primary tissues,
CC   to regenerate tissues, to identify their cognate ligands or binding
CC   partners, and in chemotaxis, and can be used as a food additive or
CC   preservative to modify storage properties. Antibodies specific for a
CC   protein of the invention can be used in alleviating symptoms associated
CC   with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC   radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC   The present sequence represents a human secreted protein variant
CC   referred to in the disclosure of the invention.
XX
SQ   Sequence   252 AA;


  Query Match             72.2%;  Score 39;  DB 22;  Length 252;
  Best Local Similarity   85.7%;  Pred. No. 46;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GLQHWVP 7
           || ||||
Db     241 qlphwvp 247


RESULT   6
AAW98891
ID   AAW98891 standard; peptide; 320 AA.
XX
AC   AAW98891;
XX
DT   05-MAY-1999  (first entry)
XX
DE   Peptide SEQ ID NO 20 from US5877281
XX
KW   Del-1; developmentally-regulated endothelial cell locus 1; cancer;
KW   discoidin I; factor VIII like domain; epidermal growth factor; EGF;
KW   diabetic retinopathy; rheumatoid arthritis; endometriosis;
KW   angiogenesis.
XX
OS   Unidentified.
XX
PN   US5877281-A.
XX
PD   02-MAR-1999.
XX
PF   05-JUN-1996;   96US-0659235.
XX
PR   05-JUN-1996;   96US-0659235.
PR   07-JUN-1995;   95US-0480229.
XX
PA   (PROG-) PROGENITOR INC.
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Hogan B;  Quertermous T;  Snodgrass HR;  Zupancic T;
XX
DR   WPI; 1999-189720/16.
XX
PT   Del-1 polypeptide sequences - useful for treatment of cancer,
PT   diabetic retinopathy, rheumatoid arthritis and endometriosis
XX
PS   Disclosure; Column 61-62; 73pp; English.
XX
CC   The present invention describes developmentally-regulated endothelial
CC   cell locus 1 (Del-1). The protein has epidermal growth factor (EGF) like
CC   domains and discoidin I/factor VIII-like domains. The Del-1 proteins
CC   have an inhibitory effect on angiogenesis (blood vessel growth), this
CC   activity may be useful clinically to prevent neovascularisation of
CC   tissues such as tumour nodules and prevention of metastases. The anti-
CC   angiogenic activity of Del-1 may be used to treat abnormal conditions
CC   that result from angiogenesis, including cancer, diabetic retinopathy,
CC   rheumatoid arthritis and endometriosis. Since Del-1 promotes
CC   angiogenesis it can be used to treat cardiac ischaemia, thrombotic
CC   stroke, wound healing and peripheral vascular disease. Del-1 is also
CC   useful for promoting bone formation. Del-1 binds to alpha V beta 3
CC   integrin, and is an apoptosis inducer. The present sequence represents
CC   a sequence given in the sequence listing but not mentioned further
CC   within the specification.
XX
SQ   Sequence   320 AA;


  Query Match             72.2%;  Score 39;  DB 28;  Length 320;
  Best Local Similarity   77.8%;  Pred. No. 59;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           ||| | |||
Db      28 glqrwqpel 36


RESULT   7
AAY94454
ID   AAY94454 standard; Protein; 426 AA.
XX
AC   AAY94454;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Mouse lactadherin protein.
XX
KW   Human; lactadherin; MGF-E8; anti-tumour; immune response;
KW   exosome; dendritic cell.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= Secretion_signal
FT   Protein         23..426
FT                   /label= Lactadherin
FT   Binding-site    87..89
FT                   /label= Integrin_binding_site
FT   Misc-difference 93..111
FT                   /note= "encoded by GT"
XX
PN   EP1004664-A1.
XX
PD   31-MAY-2000.
XX
PF   24-NOV-1998;   98EP-0402925.
XX
PR   24-NOV-1998;   98EP-0402925.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (CURI-) INST CURIE.
XX
DR   WPI; 2000-352547/31.
```

```
DR   N-PSDB; AAA27141.
XX
PT   Chimeric isolated (human) lactadherin polypeptide that functions as an
PT   adaptor of cross-priming to eliminate pathogenic antigens, e.g. in
PT   tumors -
XX
PS   Disclosure; Page 12; 20pp, English.
XX
CC   Lactadherin protein was found in exosomes produced by dendritic cells.
CC   The protein is involved in the phagocytosis of particulate antigens by
CC   dendritic cells. Exosomes produced by dendritic cells exposed to
CC   tumour antigens induce potent immune responses. Lactadherin or variants
CC   of it may be used in the mediation of an immune response. Variants of
CC   lactadherin may be used for inhibition and/or stimulation of the
CC   cross-priming of antigens and stimulation of the phagocytosis of
CC   antigens by dendritic cells. Compositions derived from lactadherin can
CC   also be used to monitor an immune response, more specifically a CTL
CC   (cytotoxic T-lymphocyte) response and also to produce CTLs specific
CC   for a selected antigen. The present sequence is the mouse
CC   lactadherin protein.
XX
SQ   Sequence   426 AA;


  Query Match             72.2%;  Score 39;  DB 21;  Length 426;
  Best Local Similarity   77.8%;  Pred. No. 80;
  Matches    7;  Conservative    0,  Mismatches    2,  Indels    0,  Gaps    0;

Qy       1 GLQHWVPEL 9
            ||| | |||
Db     138 glqrwqpel 146


RESULT   8
ABB39426
ID   ABB39426 standard; Peptide; 22 AA.
XX
AC   ABB39426;
XX
DT   04-FEB-2002  (first entry)
XX
DE   Peptide #6932 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234697.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Fank DR;
XX
DR   WPI; 2001-483447/52
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 27; SEQ ID No 32561; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a peptide encoded by a single exon
CC   nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   22 AA;


  Query Match             70.4%;  Score 38;  DB 22;  Length 22;
  Best Local Similarity   71.4%;  Pred. No. 5.3;
  Matches    5;  Conservative    1,  Mismatches    1,  Indels    0;  Gaps    0;

Qy       1 GLQHWVP 7
            | |||:|
Db       2 glqhwip 8


RESULT   9
ABB24200
ID   ABB24200 standard; Protein; 22 AA.
XX
AC   ABB24200;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Protein #6199 encoded by probe for measuring heart cell gene expression.
XX
KW   Human; gene expression; heart; microarray; vascular system;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   conqenital heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234697.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Fank DR;
XX
DR   WPI; 2001-489899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 15; SEQ ID No 25970; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart (see
CC   ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC   probe. The probes may be used for predicting, measuring and displaying
CC   gene expression in samples derived from the human heart via microarrays.
CC   By measuring gene expression, the probes are useful for predicting,
CC   diagnosing, grading, staging, monitoring and prognosing diseases of the
CC   human heart and vascular system e.g. cardiovascular disease,
CC   hypertension, cardiac arrhythmias and conqenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
```

```
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   22 AA;


  Query Match             70.4%;  Score 38;  DB 22;  Length 22;
  Best Local Similarity   71.4%;  Pred. No. 5.3;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GLQHWVP 7
           | |||:|
Db       2 qfqhwlp 8


RESULT  10
AAM60107
ID   AAM60107 standard; Protein; 22 AA.
XX
AC   AAM60107;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe encoded protein SEQ ID NO: 32212.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180212.
PR   26-MAY-2000; 2000US 0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
PS   Example 4; SEQ ID NO: 32212; 650pp - Sequence Listing, English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is a protein encoded by one of
CC   the probes of the invention.
XX
SQ   Sequence   22 AA;


  Query Match             70.4%;  Score 38;  DB 22;  Length 22;
  Best Local Similarity   71.4%;  Pred. No. 5.3;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GLQHWVP 7
           | |||:|
Db       2 qfqhwlp 8


RESULT  11
AAM72718
ID   AAM72718 standard; Protein; 22 AA.
XX
AC   AAM72718;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed probe encoded protein SEQ ID NO: 33024.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 33024; 658pp - Sequence Listing; English
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is a
CC   protein encoded by one of the probes of the invention.
XX
SQ   Sequence   22 AA;


  Query Match             70.4%;  Score 38;  DB 22;  Length 22;
  Best Local Similarity   71.4%;  Pred. No. 5.3;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GLQHWVP 7
           | |||:|
Db       2 qfqhwlp 8


RESULT  12
AAM19701
ID   AAM19701 standard; Protein; 22 AA.
XX
AC   AAM19701;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Peptide #6135 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
```

XX
OS Homo sapiens.
XX
PN WO200157278 A2.
XX
PD 09-AUG-2001
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR,
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 27; SEQ ID No 24527; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP; see AAI10668-AAI28459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 22 AA;

Query Match 70.4%; Score 38; DB 22; Length 22;
Best Local Similarity 71.4%; Pred. No. 5.3;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GLQHWVP 7
| ||||:|
Db 2 gfqhwlp 8

RESULT 13
AAM32949
ID AAM32949 standard, Protein, 22 AA.
XX
AC AAM32949;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #6986 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray, human, placenta, antenatal diagnosis,
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272 A2.
XX
PD 09-AUG-2001
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR,
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 27; SEQ ID No 33218; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP;
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
SQ Sequence 22 AA;

Query Match 70.4%; Score 38; DB 22; Length 22;
Best Local Similarity 71.4%; Pred. No. 5.3;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GLQHWVP 7
| ||||:|
Db 2 gfqhwlp 8

RESULT 14
ABB68244
ID ABB68244 standard, Protein; 665 AA.
XX
AC ABB68244;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 31524.
XX
KW Drosophila, developmental biology; cell signalling, insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042 A2
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL12347.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell cell
PT interactions -
XX

PS Disclosure; SEQ ID NO 31524; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL40511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 665 AA;

Query Match 70.4%; Score 38; DB 22; Length 665;
Best Local Similarity 71.4%; Pred. No. 1.9e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GLQHWVP 7
||:|| |
Db 33 qlehwtp 39

RESULT 15
AAE13839
ID AAE13839 standard; Protein; 4019 AA.
XX
AC AAE13839;
XX
DT 26-FEB-2002 (first entry)
XX
DE Human lung tumour-specific protein SCC2-29.
XX
KW Human; lung tumour protein, immunostimulant; cytostatic; gene therapy;
KW antisense-therapy, vaccine, immune response; lung cancer, SCC2-29.
XX
OS Homo sapiens.
XX
PN WO200172295-A2.
XX
PD 04-OCT-2001.
XX
PF 28-MAR-2001; 2001WO-US09991.
XX
PR 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indillas CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX
DR WPI; 2001-639201/73.
DR N-PSDB; AAD23451.
XX
PT New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e g lung cancer -
XX
PS Disclosure; Page 309-318; 378pp; English.
XX
CC The invention relates to isolated lung tumour specific proteins and
CC their corresponding cDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is human lung tumour-specific protein.
XX
SQ Sequence 4019 AA;

Query Match 70.4%; Score 38; DB 22; Length 4019;
Best Local Similarity 85.7%; Pred. No. 1.3e+03;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LQHWVPE 8
|||| ||
Db 1475 lqhwqpe 1481

Search completed: September 5, 2002, 15:26:31
Job time: 371 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:28:52 ; Search time 69.78 Seconds
(without alignments)
12.393 Million cell updates/sec

Title: US-09-744-804-39
Perfect score: 54
Sequence: 1 GLQHWVPEL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	77.8	401	2	S65138	glycoprotein antig
2	42	77.8	409	2	T11743	pP47 protein - pig
3	42	77.8	427	2	S74211	PAS 6/7 protein pr
4	41	75.9	453	2	S74805	DNA photolyase - S
5	41	75.9	833	2	B82708	ATP dependent heli
6	40	74.1	256	2	T18787	hypothetical prot
7	40	74.1	285	2	B82336	phosphatidylserine
8	40	74.1	453	2	F96672	Similar to Flavono
9	40	74.1	480	2	T10020	hypothetical prote
10	40	74.1	512	2	G86914	probable membrane
11	40	74.1	748	2	T30634	hypothetical prote
12	39	72.2	219	2	T37794	hypothetical prote
13	39	72.2	338	2	T36025	conserved hypothet
14	39	72.2	357	2	B83205	hypothetical prote
15	39	72.2	427	2	JC4915	ags protein precur
16	39	72.2	463	1	A36479	milk fat globule m
17	39	72.2	481	2	B83062	deoxyribodipyrimid
18	39	72.2	1040	2	AD0771	probable PND-famil
19	39	72.2	1155	2	G87477	transcription repa
20	38	70.4	129	2	G83082	conserved hypothet
21	38	70.4	238	2	H87578	transcription regu
22	38	70.4	331	2	A35819	fructose-bisphosph
23	38	70.4	498	2	S68588	nicotinic acetylch
24	37	68.5	122	2	T07295	hypothetical prote
25	37	68.5	286	2	S60312	extended spectrum
26	37	68.5	448	2	C96542	hypothetical prote
27	37	68.5	992	2	T08772	hypothetical prote
28	37	68.5	1165	2	C97574	mfd protein (AF300
29	37	68.5	1165	2	AC2795	transcription-repa
30	37	68.5	1170	2	AG3595	transcription-repa
31	37	68.5	2143	2	C96595	hypothetical prote
32	37	68.5	3079	1	KGBY12	probable GTPase ac
33	37	68.5	3187	2	JC5837	364K Golgi complex
34	36	66.7	137	2	I68717	gene HLA-DQA2 prot
35	36	66.7	221	1	HLMSA1	H-2 class II histo
36	36	66.7	227	2	A47635	MHC class II histo
37	36	66.7	228	2	A53271	MHC class II histo
38	36	66.7	228	2	B53271	MHC class II histo
39	36	66.7	232	1	HLHUDQ	MHC class II histo
40	36	66.7	233	2	I79358	IA-alpha polyprote
41	36	66.7	233	2	I79357	IA-alpha polyprote
42	36	66.7	243	2	I54290	cell surface glyco
43	36	66.7	250	2	JH0749	class II histocomp
44	36	66.7	254	1	HLHUD7	MHC class II histo
45	36	66.7	254	1	HLHU3C	MHC class II histo

ALIGNMENTS

RESULT 1
S65138
glycoprotein antigen MGP57/53, mammary gland - bovine (fragment)
N;Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E
C;Species: Bos primigenius taurus (cattle)
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
C;Accession: S65138; G48394
R;Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A;Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal
A;Reference number: S65138; MUID:96125736
A;Accession: S65138
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-401 <AOK>
R;Mather, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A;Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-
II-like sequences.
A;Reference number: A48394; MUID:93250576
A;Accession: G48394
A;Status: preliminary
A;Molecule type: protein
A;Residues: 207-220 <MAT>
A;Experimental source: milk
A;Note: sequence extracted from NCBI backbone (NCBIP:131457)
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
C;Keywords: glycoprotein
F;1-32/Domain: EGF homology (fragment); EGL
F;40-79/Domain: EGF homology; EGF-
F;82-239/Domain: discoidin I amino-terminal homology; DN1
F;243-401/Domain: discoidin I amino-terminal homology; DN2

Query Match: 77.8%; Score: 42; DB 2; Length 401;
Best Local Similarity: 77.8%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GLQHWVPEL 9
||| | |||
Db 110 GLQPWAPEL 118

RESULT 2
T11743
pP47 protein - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T11743
R;Fusslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toe
Biol. Reprod. 58, 1057-1064, 1998
A;Title: Molecular cloning and characterization of P47, a novel boar sperm associated

A;Reference number: Z17325; MUID:98206817
A;Accession: T11743
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1 409 <ENS>
A;Cross-references: EMBL:Y11683, NID:g2652927, PIDN:CAA72379.1, PID:g2652928
A;Experimental source: testis
C;Function:
A;Description: may be involved in membrane remodeling and/or function as a cell pellicle
C;Superfamily: unassigned EGF-related proteins; EGF homology
F;6-40/Domain: EGF homology <EGF>

Query Match 77.8%; Score 42; DB 2; Length 409;
Best Local Similarity 77.8%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GLQHWVPEL 9
||| | |||
Db 118 GLQRWAPEL 126

RESULT 3
S74211
PAS-6/7 protein precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000
C;Accession: S74211; S78114; S24181
R;Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
Eur. J. Biochem. 240, 628-636, 1996
A;Title: Characterization of glycoprotein PAS 6/7 from membranes of bovine milk fat glob
A;Reference number: S74211; MUID:97008954
A;Accession: S74211
A;Molecule type: mRNA
A;Residues: 1 427 <HVA>
A;Cross-references: EMBL:X91895; NID:g1632778; PIDN:CAA62997.1; PID:g1632779
A;Accession: S78114
A;Molecule type: protein
A;Residues: 19-85,99-119,149-165,174-215,221-242,248-273,285-293,[illegible]
R;Kim, D.H.; Kanno, C.; Mizokami, Y.
Biochim. Biophys. Acta 1122, 203-211, 1992
A;Title: Purification and characterization of major glycoproteins, PAS 6 and PAS 7, from
A;Reference number: S23926; MUID:92353107
A;Accession: S24181
A;Molecule type: protein
A;Residues: 383-394 <KIM>
C;Superfamily: milk fat globule protein; discoidin I amino terminal homology; EGF homolo
C;Keywords: blocked amino end; disulfide bond; glycoprotein; milk
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-427/Product: PAS 6/7 protein #status experimental <MAT>
F;24-58/Domain: EGF homology <EG1>
F;66-105/Domain: EGF homology <EG2>
F;108-265/Domain: discoidin I amino terminal homology <DN1>
F;269-427/Domain: discoidin I amino terminal homology <DN2>
F;24-45,29-47,49-58,65-77,71-94,96-105/Disulfide bonds: #status predicted
F;27/Binding site: carbohydrate (Ser) (covalent) #status experimental
F;34/Binding site: carbohydrate (Thr) (covalent) #status experimental
F;59,227/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;109-265,252-256,270-427/Disulfide bonds: #status experimental

Query Match 77.8%; Score 42; DB 2; Length 427;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GLQHWVPEL 9
||| | |||
Db 136 GLQRWAPEL 144

RESULT 4
S74805
DNA photolyase - Synechocystis sp. (strain PCC 6803)
N;Alternate names: protein slll629
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S74805
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys s.
A;Reference number: S74322; MUID:97061201
A;Accession: S74805
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-453 <KAN>
A;Cross-references: EMBL:D90909; GB:AB001339; NID:g1652841; PIDN:BAA17766.1; PID:g165
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Gene: phr
A;Start codon: GTG
C;Superfamily: deoxyribodipyrimidine photo-lyase

Query Match 75.9%; Score 41; DB 2; Length 454;
Best Local Similarity 75.0%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LQHWVPEL 9
|:|||:||
Db 392 IPHWIPEL 399

RESULT 5
E82708
ATP-dependent helicase XF1229 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000
C;Accession: E82708
R;Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: E82708
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-833 <SIM>
A;Cross-references: GB:AE003956; GB:AE003849; NID:g9106196; PIDN:AAF84039.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R. Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer as Neto, H.; Cerqueira, G.C.; Cruz, K.S.; Dias-Neto, E.; Eberlin, M.N.; Fagen, J.R. submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr J.P.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins A.A.thors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C. , P.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1229
C;Superfamily: ATP-dependent RNA helicase hrpB

Query Match 75.9%; Score 41; DB 2; Length 833;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LQHWVPEL 9

```
          |:|||:|||
Db    658 LRHWMPEL 665


RESULT   6
T48787
hypothetical protein 13E11.350 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 28-Jul-2000
C;Accession: T48787
R;Schulte, U ; Aign, V ; Hoheisel, J.; Brandt, P.; Partmann, B., Holland, P ., Nyakatura,
submitted to the Protein Sequence Database, April 2000
A;Reference number: 224541
A;Accession: T48787
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-256 <SCH>
A;Cross-references: EMBL:AL353820; GSPDB:GN00112; NCSP:13E11.350
A;Experimental source: cosmid contig 13E11; strain 74
C;Genetics:
A;Gene: NCSP:13E11.350
A;Map position: 2
A;Introns: 213/2
C;Superfamily: Neurospora crassa hypothetical protein 13E11.350


  Query Match             74.1%;  Score 40;  DB 2;  Length 256;
  Best Local Similarity   66.7%;  Pred. No. 14;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           ||: ||||:
Db     141 GLEEWVPEV 149


RESULT   7
B82336
phosphatidylserine decarboxylase (EC 4.1.1.65) [similarity] - Vibrio cholerae (strain N
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 24-Aug-2001
C;Accession: B82336
R;Heidelberg, J.F ; Eisen, J.A.; Nelson, W.C ; Clayton, R.A ; Gwinn, M.L.; Dodson, R.J ;
chardson, D.; Ermolaeva, M.D., Vamathevan, J., Bass, S.; Qin, H., Dragoi, I ; Sellers, P
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae
A;Reference number: A82035; MUID:20406833
A;Accession: B82336
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-285 <HEI>
A;Cross-references: GB:AE004122; GB:AE003852; NID:g9654745; PIDN:AAF93512.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961, biotype El Tor
C;Genetics:
A;Gene: VC0339
A;Map position: 1
C;Superfamily: Escherichia coli phosphatidylserine decarboxylase
C;Keywords: blocked amino end; carbon-carbon lyase; carboxy-lyase
F;251-252/Cleavage site: Gly-Ser (autolytic) #status predicted
F;252/Modified site: pyruvic acid (Ser) (in mature form) #status predicted


  Query Match             74.1%;  Score 40;  DB 2;  Length 285;
  Best Local Similarity   62.5%;  Pred. No. 16;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPE 8
           |||:|:|:
Db       7 GLQYWIPQ 14


RESULT   8
F96672
Similar to Flavonol 3-O-Glucosyltransferase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: F96672
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A., Kaul, S., White, O.; Alon
 Chin, C.W.; Ching, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
anson, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, A.; Jenkins, T.; Johnson-Hopson, C.; Khan, S.; Khaykin, E., Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Fiuo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: F96672
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-452 <STO>
A;Cross-references: GB:AF065174; NID:g5042427; PIDN:AAD08250.1; GSPDB:GN00141
C;Genetics:
A;Gene: F13O11.22
A;Map position: 1


  Query Match             74.1%;  Score 40;  DB 2;  Length 452;
  Best Local Similarity   75.0%;  Pred. No. 25;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
            | |||||:
Db     115 LAHWVPEM 122


RESULT   9
T10029
hypothetical protein MLCB628.10c - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
C;Accession: T10029
R;Eiglmeier, K.; Honore, N., Woods, S.A.; Caudron, B., Cole, S.T.
Mol. Microbiol. 7, 197-206, 1993
A;Title: Use of an ordered cosmid library to deduce the genomic organization of Mycob
A;Reference number: Z16917; MUID:94188700
A;Accession: T10029
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-480 <EIG>
A;Cross-references: EMBL:Y14967; NID:g2370268; PIDN:CAA75138.1; PID:g2370277
C;Genetics:
A;Note: MLCB628.10c


  Query Match             74.1%;  Score 40;  DB 2;  Length 480;
  Best Local Similarity   85.7%;  Pred. No. 27;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GLQHWVP 7
           | |||||
Db     247 GYQHWVP 253


RESULT  10
G86914
probable membrane protein [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 27-Apr-2001
C;Accession: G86914
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holro
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Buttet, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.;
```

A;Title: Massive gene decay in the leprosy bacillus
A;Reference number: A86099; MUID:21287742; PMID:11234002
A;Accession: G86914
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-512 <STO>
A;Cross-references: GB:AL450380; NID:g13092444; PIDN:CAC29555.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML0047

Query Match 74.1%; Score 40; DB 2; Length 512;
Best Local Similarity 85.7%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GLQHWVP 7
| ||||||
Db 279 GYQHWVP 285

RESULT 11
T30634
hypothetical protein 32L - Molluscum contagiosum virus 1
N;Alternate names: MC032L
C;Species: Molluscum contagiosum virus 1
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jul-2000
C;Accession: T30634
R;Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.
Science 273, 813-816, 1996
A;Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re
A;Reference number: Z20876; MUID:96325459
A;Accession: T30634
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-748 <SEN>
A;Cross-references: EMBL:U60315; NID:g1491343; PIDN:AAC55160.1; PID:g1491975
C;Genetics:
A;Note: MC032L
C;Superfamily: vaccinia virus hypothetical protein 49L

Query Match 74.1%; Score 40; DB 2; Length 748;
Best Local Similarity 85.7%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 QHWVPEL 9
:||||||
Db 52 RHWVPEL 58

RESULT 12
T37794
hypothetical protein SPAC16E8.14c - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
C;Accession: T37794
R;Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1995
A;Reference number: Z21746
A;Accession: T37794
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-219 <OLI>
A;Cross-references: EMBL:Z98529; PIDN:CAB11642.1; GSPDB:GN00066; SPDB:SPAC16E8.14c
A;Experimental source: strain 972h-, cosmid c16E8
C;Genetics:
A;Gene: SPDB:SPAC16E8.14c
A;Map position: 1
A;Introns: 12/2; 32/3; 122/1; 158/3
C;Superfamily: Neurospora crassa hypothetical protein 13E11.350

Query Match 72.2%; Score 39; DB 2; Length 219;
Best Local Similarity 75.0%; Pred. No. 18;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GLQHWVPE 8
|||:| ||
Db 110 GLQNWTPE 117

RESULT 13
T36025
conserved hypothetical protein SCC54.12c - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 19-May-2000
C;Accession: T36025
R;Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z21581
A;Accession: T36025
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-338 <SEE>
A;Cross-references: EMBL:AL035591; PIDN:CAB38142.1; GSPDB:GN00070; SCOEDB:SCC54.12c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCC54.12c
C;Superfamily: Bacillus subtilis conserved hypothetical protein yvcJ

Query Match 72.2%; Score 39; DB 2; Length 338;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 HWVPEL 9
||||||
Db 246 HWVPEL 251

RESULT 14
E83205
hypothetical protein PA3515 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: E83205
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa
A;Reference number: A82950; MUID:20437337
A;Accession: E83205
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-357 <STO>
A;Cross-references: GB:AE004772; GB:AE004091; NID:g9946062; PIDN:AAG06003.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3515

Query Match 72.2%; Score 39; DB 2; Length 357;
Best Local Similarity 62.5%; Pred. No. 29;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LQHWVPEL 9
.|||:| |
Db 160 VQHWIPSL 167

RESULT 15
JC4915
ags protein precursor - rat
N;Alternate names: O acetyl-Gd3 ganglioside
C;Species: Rattus norvegicus (Norway rat)

C;Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
C;Accession: JC4915
R;Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A;Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
A;Reference number: JC4915; MUID:96374422
A;Accession: JC4915
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-427 <OGU>
A;Cross-references: DDBJ:D84066; NID:g1620906; PIDN:BAA12210.1; PID:g1620907
A;Experimental source: CST cell
C;Comment: This protein is required for the O-acetylation of disialoganglioside sialic a
C;Genetics:
A;Gene: ags
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
F;1-21/Domain: signal sequence #status predicted <SIG>
F;28-60/Domain: EGF homology <EGF1>
F;68-107/Domain: EGF homology <EGF2>
F;110-267/Domain: discoidin I amino-terminal homology <DIN1>
F;271-427/Domain: discoidin I amino-terminal homology <DIN2>

Query Match 72.2%; Score 39; DB 2, Length 427;
Best Local Similarity 77.8%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GLQHWVPEL 9
|||| | |||
Db 138 GIQRWGPEL 146

Search completed: September 5, 2002, 15:28:53
Job time: 358 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:42 ; Search time 22.99 Seconds
(without alignments)
10.252 Million cell updates/sec

Title: US-09-744-804-39
Perfect score: 54
Sequence: 1 GLQHWVPEL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	54	100.0	387	1	MFGM_HUMAN	Q08431 homo sapien
2	42	77.8	409	1	MFGM_PIG	P79385 sus scrofa
3	42	77.8	427	1	MFGM_BOVIN	Q95114 bos taurus
4	39	72.2	219	1	YDRE_SCHPO	Q13748 schizosacch
5	39	72.2	427	1	MFGM_RAT	P70492 rattus norv
6	39	72.2	463	1	MFGM_MOUSE	P21956 mus musculu
7	38	70.4	331	1	F16P_PHOSH	P22780 rhodobacter
8	38	70.4	498	1	ACH1_CAEEL	P48180 caenorhabdi
9	37	68.5	4079	1	IPAJ_YEAST	P19158 saccharomyc
10	36	66.7	221	1	HA2Q_MOUSE	P04227 mus musculu
11	36	66.7	227	1	HA2U_MOUSE	P14438 mus musculu
12	36	66.7	232	1	HA24_HUMAN	P04225 homo sapien
13	36	66.7	233	1	HA2F_MOUSE	P14435 mus musculu
14	36	66.7	233	1	HA2P_MOUSE	P14436 mus musculu
15	36	66.7	233	1	HA2S_MOUSE	P14437 mus musculu
16	36	66.7	254	1	HA22_HUMAN	P04226 homo sapien
17	36	66.7	254	1	HA23_HUMAN	P01909 homo sapien
18	36	66.7	254	1	HA27_HUMAN	P05536 homo sapien
19	36	66.7	254	1	HA2C_PIG	P15980 sus scrofa
20	36	66.7	254	1	HA2J_MOUSE	P23150 mus musculu
21	36	66.7	255	1	HA21_HUMAN	P01908 homo sapien
22	36	66.7	255	1	HA25_HUMAN	P01907 homo sapien
23	36	66.7	255	1	HA26_HUMAN	P01906 homo sapien
24	36	66.7	255	1	HA2D_PIG	P15981 sus scrofa
25	36	66.7	256	1	HA2B_MOUSE	P14434 mus musculu
26	36	66.7	256	1	HA2B_RAT	P20037 rattus norv
27	36	66.7	256	1	HA2D_MOUSE	P04228 mus musculu
28	36	66.7	256	1	HA2K_MOUSE	P01910 mus musculu
29	36	66.7	529	1	CN1A_BOVIN	P14103 bos taurus
30	36	66.7	534	1	CN1A_HUMAN	P54750 homo sapien
31	36	66.7	557	1	COX1_NEUCR	P03945 neurospora
32	36	66.7	565	1	CN1A_MOUSE	Q61481 mus musculu
33	36	66.7	1101	1	YA4C_SCHPO	Q09733 schizosacch
34	36	66.7	1456	1	RRPO_PVX	P09395 potato viru
35	36	66.7	1456	1	RRPO_PVXX4	P17775 potato viru
36	35	64.8	261	1	1AM_MYCTU	O53698 mycobacteri
37	35	64.8	300	1	ICIA_PSEAE	Q9hw38 pseudomonas
38	35	64.8	517	1	COX1_MYXGL	O21079 myxine glut
39	35	64.8	552	1	COX1_SYNVU	P50676 synechococc
40	35	64.8	571	1	SYE_PYRAB	Q9y1e3 pyrococcus
41	35	64.8	836	1	CCS1_HUMAN	Q13724 homo sapien
42	35	64.8	1043	1	SYI_CHLPN	Q92972 chlamydia p
43	34	63.0	146	1	AP4A_HUMAN	P53585 homo sapien
44	34	63.0	146	1	AP4A_MOUSE	P56380 mus musculu
45	34	63.0	146	1	AP4A_PIG	P50584 sus scrofa

ALIGNMENTS

RESULT 1
MFGM_HUMAN
ID MFGM_HUMAN STANDARD; PRT; 387 AA.
AC Q08431;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule EGF factor 8) (MFG-E8) (HMFG)
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
GN MFGE8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast, and Breast carcinoma;
RX MEDLINE=96213908; PubMed=8639264;
RA Couto J.R., Taylor M.R., Godwin S.G., Coriani R.I., Peterson J.A.;
RT "Cloning and sequence analysis of human breast epithelial antigen
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal
RT growth factor-like domain.";
RL DNA Cell Biol. 15:281-286(1996).
RN [2]
RP SEQUENCE OF 170-387 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=91371351; PubMed=1909932;
RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
RA Coriani R.L.;
RT "A Mr 46,000 human milk fat globule protein that is highly expressed
RT in human breast tumors contains factor VIII-like domains.";
RL Cancer Res. 51:4994-4998(1991).
RN [3]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=98194921; PubMed=9535276;
RA Giuffrida M.G., Caroletto M., Giunta C., Conti A.,
RA Godovac-Zimmermann J.;
RT "Isolation and characterization of full and truncated forms of human
RT breast carcinoma protein BA46 from human milk fat globule membranes.";
RL J. Protein Chem. 17:143-148(1998).
RN [4]
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
RX MEDLINE=99342076; PubMed=10411933;
RA Haggqvist B., Naslund J., Sletten K., Westermark G.T., Mucchiano G.,
RA Tjernberg L.O., Nordstedt C., Engstrom U., Westermark P.;
RT "Medin: an integral fragment of aortic smooth muscle cell-produced
RT lactadherin forms the most common human amyloid.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=97405885; PubMed=9260929;
RA Taylor M.R., Couto J.R., Scallan C.D., Coriani R.L., Peterson J.A.;
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
RT (RGD)-dependent cell adhesion.";

```
RL   DNA Cell Biol. 16:861-869(1997).
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
CC       SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
CC   -!- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
CC       MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC   -!- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
CC       AMINO ACID 264 AND 273.
CC   -!- SIMILARITY: CONTAINS 1 EGF LIKE DOMAIN
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------- ----------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U58516; AAC50549.1; -.
DR   EMBL; S56151; AAB19771.1; -.
DR   MIM; 602281; -.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
FT   SIGNAL       1     23
FT   CHAIN       24    387       LACTADHERIN.
FT   CHAIN      202    387       LACTADHERIN, SHORT FORM.
FT   CHAIN      208    317       MEDIN.
FT   DOMAIN      24     67       EGF-LIKE.
FT   DOMAIN      70    225       F5/8 TYPE C 1.
FT   DOMAIN     230    387       F5/8 TYPE C 2.
FT   SITE        46     48       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    27     38       BY SIMILARITY.
FT   DISULFID    32     55       BY SIMILARITY.
FT   DISULFID    57     66       BY SIMILARITY.
FT   DISULFID    70    225       BY SIMILARITY.
FT   DISULFID   212    216       BY SIMILARITY.
FT   DISULFID   230    387       BY SIMILARITY.
FT   CARBOHYD   238    238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   325    325       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   329    329       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   350    350       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   387 AA;  42123 MW;  2EB657[illegible] CRC64;


  Query Match             100.0%;  Score 54;  DB 1;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.019;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           |||||||||
Db      97 GLQHWVPEL 105


RESULT   2
MFGM_PIG
ID   MFGM_PIG       STANDARD;      PRT;   409 AA.
AC   P79385;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Lactadherin (Milk fat globule EGF factor 8) (MFG-E8) (MFGM) (Sperm
DE   surface protein SP47) (PP47).
```

```
GN   MFGE8.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION. MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC       SPERMATOZOAN.
CC   -!- SIMILARITY: CONTAINS 2 EGF LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Y11683; CAA72379.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Glycoprotein, Repeat, EGF like domain.
FT   DOMAIN       2     41       EGF-LIKE 1.
FT   DOMAIN      44     88       EGF-LIKE 2.
FT   DOMAIN      91    247       F5/8 TYPE C 1.
FT   DOMAIN     252    409       F5/8 TYPE C 2.
FT   SITE        67     69       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     6     17       BY SIMILARITY.
FT   DISULFID    11     29       BY SIMILARITY.
FT   DISULFID    31     40       BY SIMILARITY.
FT   DISULFID    91    247       BY SIMILARITY.
FT   DISULFID   234    238       BY SIMILARITY.
FT   DISULFID   252    409       BY SIMILARITY.
FT   CARBOHYD    41     41       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   372    372       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   409 AA;  45735 MW;  [illegible]7AFB0329927A CRC64;


  Query Match             77.8%;  Score 42;  DB 1;  Length 409;
  Best Local Similarity   77.8%;  Pred. No. 2.8;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           ||| | |||
Db     118 GLQRWAPEL 126


RESULT   3
MFGM_BOVIN
ID   MFGM_BOVIN     STANDARD;      PRT;   427 AA.
AC   Q95114; Q27959; P79344;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
DE   (MGF57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
```

```
DE   SP47) (BP47) (Components 15/16).
GN   MFGE8.
OS   Bos taurus (Bovine)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC   STRAIN=HOLSTEIN; TISSUE=Mammary gland;
RX   MEDLINE=97008954; PubMed=8856064;
RA   Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RA   Petersen T.E.;
RT   "Characterization of glycoprotein PAS 6/7 from membranes of bovine
RT   milk fat globules.";
RL   Eur. J. Biochem. 240:628-636(1996).
RN   [2]
RP   SEQUENCE OF 18-427 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=96125736; PubMed=8541316;
RA   Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura E.,
RA   Matsuda T.;
RT   "Molecular cloning of glycoprotein antigens MG1/3/5 to 8 reactive with
RT   monoclonal antibodies raised against bovine milk fat globule
RT   membrane.";
RL   Biochim. Biophys. Acta 1245:385-391(1995).
RN   [3]
RP   SEQUENCE OF 19-427 FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
RC   TISSUE=Milk;
RX   MEDLINE=92250576; PubMed=8485470;
RA   Mather I.H., Banghart L.R., Lane W.S.;
RT   "The major fat-globule membrane proteins, bovine components 15/16 and
RT   guinea-pig GP 55, are homologous to MFG-E8, a murine glycoprotein
RT   containing epidermal growth factor-like and factor V/VIII-like
RT   sequences.";
RL   Biochem. Mol. Biol. Int. 29:545-554(1993).
CC   -!- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
CC       MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM
CC       LACKS 53 AMINO ACIDS WITHIN THE F5/8 TYPE C 1 DOMAIN.
CC   -!- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
CC   -!- PTM: THE 2,6-LINKED GLYCANS CONSTIST OF GAL, GLCNAC AND FUC, WITH
CC       PROBLABLY FUC AS REDUCING TERMINAL SUGAR.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X91895; CAA62997.1; -.
DR   EMBL; S80643; AAB35894.2; -.
DR   EMBL; Y11719; CAA72406.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain;
KW   Alternative splicing.
FT   SIGNAL        1     18
FT   CHAIN        19    427       LACTADHERIN.
FT   DOMAIN       20     59       EGF-LIKE 1.
FT   DOMAIN       62    106       EGF-LIKE 2.
FT   DOMAIN      109    265       F5/8 TYPE C 1.
FT   DOMAIN      270    427       F5/8 TYPE C 2.
FT   SITE         85     87       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     24     35       BY SIMILARITY.
FT   DISULFID     29     47       BY SIMILARITY.
FT   DISULFID     49     58       BY SIMILARITY.
FT   DISULFID     66     77       BY SIMILARITY.
FT   DISULFID     71     94       BY SIMILARITY.
FT   DISULFID     96    105       BY SIMILARITY.
FT   DISULFID    109    265
FT   DISULFID    252    256
FT   DISULFID    270    427
FT   CARBOHYD     27     27       O-LINKED (FUC, ...) (IN PAS-6).
FT   CARBOHYD     34     34       O-LINKED (FUC, ...) (IN PAS-7).
FT   CARBOHYD     53     53       N-LINKED (GLCNAC, ...) (BY SIMILARITY) (IN PAS-6
FT                                AND PAS-7).
FT   CARBOHYD    227    227       N-LINKED (GLCNAC, ...) (HIGH MANNOSE) (IN
FT                                PAS-6).
FT   VARSPLIC    169    221       MISSING (IN SHORT ISOFORM).
FT   CONFLICT     19     19       A -> F (IN REF. 1).
FT   CONFLICT     28     28       L -> Q (IN REF. 1).
SQ   SEQUENCE   427 AA;  47411 MW;  4CBBEE3A1DC4EB24 CRC64;


  Query Match             77.8%;  Score 42;  DB 1;  Length 427;
  Best Local Similarity   77.8%;  Pred. No. 2.9;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           ||| | |||
Db     136 GLQRWAPEL 144


RESULT   4
YDRE_SCHPO
ID   YDRE_SCHPO     STANDARD;      PRT;   219 AA.
AC   O13748;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Hypothetical 24.4 kDa protein C16L8.14C in chromosome I.
GN   SPAC16E8.14C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Olivo R., Barrio R., Barrell B.G., Rajandream M.A., Wood V.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO YEAST YBR261C.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z98529; CAB11042.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   219 AA;  24388 MW;  3CBEF538C0C99C44 CRC64;
```

```
  Query Match             72.2%;  Score 39;  DB 1;  Length 219;
  Best Local Similarity   75.0%;  Pred. No. 5;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPE 8
            |||:| ||
Db     110 GLQNWTPE 117


RESULT   5
MFGM_RAT
ID   MFGM_RAT        STANDARD;      PRT;   427 AA.
AC   P70490;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (O-
DE   acetyl GD3 ganglioside synthase) (AGS) (MFGM)
GN   MFGE8 OR AGS.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96374422; PubMed=8780713;
RA   Oqura K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
RT   "Cloning and expression of cDNA for O-acetylation of GD3
RT   ganglioside.";
RL   Biochem. Biophys. Res. Commun. 225:932-938(1996).
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO
CC       PARTICIPATE IN THE O-ACETYLATION OF GD3 GANGLIOSIDE SIALIC ACID.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D84068; BAA12210.1; -.
DR   HSSP; P00740; 1IXA.
DR   InterPro; IPR000561; EGF-like
DR   InterPro; IPR001438; EGF_11.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   PRINTS; PR00010; EGFBLOOD
DR   SMART; SM00181; EGF; 2
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Repeat; EGF-like domain; Milk
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23    427       LACTADHERIN.
FT   DOMAIN      24     61       EGF-LIKE 1
FT   DOMAIN      64    108       EGF-LIKE 2
FT   DOMAIN     111    267       F5/8 TYPE C 1.
FT   DOMAIN     272    427       F5/8 TYPE C 2.
FT   DISULFID    28     39       BY SIMILARITY.
FT   DISULFID    33     49       BY SIMILARITY.
FT   DISULFID    51     60       BY SIMILARITY
FT   DISULFID    68     79       BY SIMILARITY.
FT   DISULFID    73     96       BY SIMILARITY.
FT   DISULFID    98    107       BY SIMILARITY.
FT   DISULFID   111    267       BY SIMILARITY.
FT   DISULFID   254    258       BY SIMILARITY.
FT   DISULFID   272    427       BY SIMILARITY
FT   SITE        87     89       CELL ATTACHMENT SITE (POTENTIAL).
FT   CARBOHYD    61     61       N-LINKED (GLCNAC      ) (POTENTIAL).
FT   CARBOHYD   230    230       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   280    280       N-LINKED (GLCNAC    .) (POTENTIAL).
FT   CARBOHYD   390    390       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   427 AA;  47413 MW;  EA8C8631F3EE6047 CRC64;


  Query Match             72.2%;  Score 39;  DB 1;  Length 427;
  Best Local Similarity   77.8%;  Pred. No. 9.9;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
            ||| | |||
Db     138 GLQRWGPEL 146


RESULT   6
MFGM_MOUSE
ID   MFGM_MOUSE      STANDARD;      PRT;   463 AA.
AC   P21956; P97800;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM)
DE   (Sperm surface protein SP47) (MP47)
GN   MFGE8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.
RC   TISSUE=Mammary gland;
RX   MEDLINE=91048008; PubMed=2122462;
RA   Stubbs J.D., Lekutis C., Singer K.L., Bui A., Yuzuki D.,
RA   Srinivasan U., Parry G.;
RT   "cDNA cloning of a mouse mammary epithelial cell surface protein
RT   reveals the existence of epidermal growth factor-like domains linked
RT   to factor VIII-like sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
RN   [2]
RP   SEQUENCE OF 23-463 FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC       SPERMATOZOAN.
CC   -!- DEVELOPMENTAL STAGE: MRNA EXPRESSION IS DETECTABLE IN MAMMARY
CC       TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M38337; AAA39534.1; -.
DR   EMBL; Y11684; CAA72380.1; -.
DR   PIR; A36479; A36479
DR   HSSP; P00740; 1EDM.
```

```
DR   MGD; MGI:102768; Mfge8.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Repeat; EGF like domain; Milk.
FT   SIGNAL       1     22
FT   CHAIN       23    463       LACTADHERIN
FT   DOMAIN      24     61       EGF-LIKE 1.
FT   DOMAIN      64    108       EGF LIKE 2.
FT   DOMAIN     148    303       F5/8 TYPE C 1.
FT   DOMAIN     308    463       F5/8 TYPE C 2.
FT   SITE        87     89       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    28     39       BY SIMILARITY.
FT   DISULFID    33     49       BY SIMILARITY
FT   DISULFID    51     60       BY SIMILARITY.
FT   DISULFID    68     79       BY SIMILARITY.
FT   DISULFID    73     96       BY SIMILARITY.
FT   DISULFID    98    107       BY SIMILARITY
FT   DISULFID   148    303       BY SIMILARITY
FT   DISULFID   290    294       BY SIMILARITY
FT   DISULFID   308    463       BY SIMILARITY.
FT   CARBOHYD    61     61       N-LINKED (GLCNAC...) (POTENTIAL)
FT   CARBOHYD   266    266       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   316    316       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   426    426       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT    30     30       S -> F (IN REF. 2).
FT   CONFLICT    35     35       N -> D (IN REF. 1, AA SEQUENCE)
FT   CONFLICT   110    147       ETNYYNLDGEYMFTTAVPNTAVPTPAPTPDLSNNLASR ->
FT                               G (IN REF. 2).
FT   CONFLICT   168    168       Y -> S (IN REF. 2).
FT   CONFLICT   196    196       H -> T (IN REF. 2).
FT   CONFLICT   309    309       L -> S (IN REF. 2).
FT   CONFLICT   395    395       E -> A (IN REF. 2).
SQ   SEQUENCE   463 AA;  51465 MW;  D7BB6C6EFBBA724D CRC64;


  Query Match             72.2%;  Score 39;  DB 1;  Length 463;
  Best Local Similarity   77.8%;  Pred. No. 11;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           |:| | |||
Db     175 GLQRWGPEI 183


RESULT   7
F16R_RHOSH
ID   F16R_RHOSH     STANDARD;      PRT;   331 AA.
AC   P22780;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Fructose-1,6-bisphosphatase II (EC 3.1.3.11) (D-fructose-1,6-
DE   bisphosphate 1-phosphohydrolase) (FBPase II).
GN   FBPB.
OS   Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1063;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91084479; PubMed=2175647;
RA   Gibson J.L., Chen L.-H., Tower P.A., Tabita F.R.;
RT   "The form II fructose 1,6-bisphosphatase and phosphoribulokinase
RT   genes form part of a large operon in Rhodobacter sphaeroides: primary
RT   structure and insertional mutagenesis analysis.";
RL   Biochemistry 29:8085-8093(1990).
CC   -!- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate + H(2)O = D-
CC       fructose 6-phosphate + phosphate.
CC   -!- ENZYME REGULATION: FBPase II,6-BISPHOSPHATASE II IS NOT LIGHT-
CC       ACTIVATED.
CC   -!- PATHWAY: KEY ENZYME OF THE REDUCTIVE PENTOSE PHOSPHATE PATHWAY
CC       OR CALVIN CYCLE OF PHOTOSYNTHETIC CARBON DIOXIDE ASSIMILATION.
CC   -!- MISCELLANEOUS: THERE ARE TWO GENES FOR FBPASE IN R.SPHAEROIDES.
CC   -!- SIMILARITY: BELONGS TO THE FBPASE FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; J02922; AAA26105.1; -.
DR   PIR; A35819; A35819.
DR   HSSP; P22419; 1SPI.
DR   InterPro; IPR000146; In_FB_phphtase.
DR   Pfam; PF00316; FBPase; 1.
DR   PRINTS; PR00377; INFBPHPHTASE.
DR   ProDom; PD001491; In_FB_phphtase; 1.
DR   PROSITE; PS00124; FBPASE; 1.
KW   Hydrolase; Carbohydrate metabolism; Calvin cycle; Photosynthesis;
KW   Multigene family.
FT   ACT_SITE    255    255       BY SIMILARITY
SQ   SEQUENCE   331 AA;  35250 MW;  509EB62E9BBEC013 CRC64;


  Query Match             70.4%;  Score 38;  DB 1;  Length 331;
  Best Local Similarity   66.7%;  Pred. No. 11;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           |:|||| :|
Db     157 GVQHWVLDL 165


RESULT   8
ACH1_CAEEL
ID   ACH1_CAEEL     STANDARD;      PRT;   498 AA.
AC   P48180;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Acetylcholine receptor like protein, alpha type chain precursor.
GN   F25G6.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=96136476; PubMed=8627624;
RA   Baillivet M., Allioli C., Bertrand S., Bertrand D.;
RT   "Novel nAChR acetylcholine receptors in the nematode Caenorhabditis
RT   elegans.";
RL   J. Mol. Biol. 258:261-269(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Nelson J., Wohldmann P.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POSSIBLE ACETYLCHOLINE RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE LIGAND-GATED ION CHANNELS FAMILY.
CC   ---------------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   -----------------------------------------------------------------------
DR   EMBL; X83887; CAA58764.1; -.
DR   EMBL; AF022973; AAC25796.1; -.
DR   WormPep; F25G6.3; CE09639
DR   InterPro; IPR000188; GABAA_receptor.
DR   InterPro; IPR001175; Neur_channel.
DR   Pfam; PF02931; Neur_chan_LBD; 1.
DR   Pfam; PF02932; Neur_chan_memb; 1.
DR   PRINTS; PR00252; NRIONCHANNEL
DR   PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
KW   Receptor; Postsynaptic membrane; Ionic channel; Glycoprotein;
KW   Transmembrane; Signal
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    498       ACETYLCHOLINE RECEPTOR LIKE PROTEIN,
FT                                ALPHA-TYPE CHAIN.
FT   DOMAIN       20    230       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    231    252       POTENTIAL.
FT   TRANSMEM    261    279       POTENTIAL.
FT   TRANSMEM    295    314       POTENTIAL.
FT   DOMAIN      315    472       CYTOPLASMIC (POTENTIAL)
FT   TRANSMEM    473    493       POTENTIAL.
FT   DISULFID    147    161       BY SIMILARITY.
FT   DISULFID    211    212       ASSOCIATED WITH RECEPTOR ACTIVATION
FT                                (BY SIMILARITY).
FT   CARBOHYD     43     43       N-LINKED (GLCNAC . .) (POTENTIAL).
FT   CARBOHYD     93     93       N-LINKED (GLCNAC ...) (POTENTIAL).
SQ   SEQUENCE   498 AA;  57169 MW;  F4B3ABB40AC9FAB2 CRC64;


  Query Match             70.4%;  Score 38;  DB 1;  Length 498;
  Best Local Similarity   85.7%;  Pred. No. 17;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GLQHWVP 7
           || ||||
Db     133 GLVHWVP 139


RESULT   9
IRA2_YEAST
ID   IRA2_YEAST     STANDARD;      PRT;  3079 AA.
AC   P19158; Q08239;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Inhibitory regulater protein IRA2.
GN   IRA2 OR CIC4 OR CCS1 OR YeL068W OR G0985
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90318397; PubMed=2164637;
RA   Tanaka K., Nakafuku M., Tamanoi F., Kaziro Y., Matsumoto K., Toh-E A.;
RT   "IRA2, a second gene of Saccharomyces cerevisiae that encodes a
RT   protein with a domain homologous to mammalian ras GTPase-activating
RT   protein.";
RL   Mol. Cell. Biol. 10:4303-4313(1990).
RN   [2]
RP   SEQUENCE OF 1-2423 FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE=95208358; PubMed=7900427;
RA   Zumstein E., Griffin H., Schweizer M.,
RT   "Sequence of a 10.27 kb segment on the left arm of chromosome XV from
RT   Saccharomyces cerevisiae includes part of the IRA2 gene and a
RT   putative new gene.";
RL   Yeast 10:1583-1587(1994).
RN   [3]
RP   SEQUENCE OF 1982-3079 FROM N.A.
RX   MEDLINE=97521807; PubMed=9178509;
RA   Izermia M., Katsoulou C., Alexandraki D.;
RT   "Sequence analysis of a 33.2 kb segment from the left arm of yeast
RT   chromosome XV reveals eight known genes and ten new open reading
RT   frames including homologues of ABC transporters, inositol
RT   phosphatases and human expressed sequence tags.";
RL   Yeast 13:583-589(1997).
RN   [4]
RP   IDENTIFICATION OF CCS1 AS IRA2.
RX   MEDLINE=92405229; PubMed=1326414;
RA   Bussereau F., Dupont C.H., Boy-Marcotte E., Mallet L., Jacquet M.;
RT   "The CCS1 gene from Saccharomyces cerevisiae which is involved in
RT   mitochondrial functions is identified as IRA2 an attenuator of RAS1
RT   and RAS2 gene products.";
RL   Curr. Genet. 21:325-329(1992).
CC   -!- FUNCTION: INHIBITORY REGULATOR OF THE RAS-CYCLIC AMP PATHWAY.
CC       STIMULATES THE GTPASE ACTIVITY OF RAS PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 RAS-GAP DOMAIN.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   -----------------------------------------------------------------------
DR   EMBL; M33779; AAA34710.1; -.
DR   EMBL; X83121; CAA58201.1; -.
DR   EMBL; X75449; CAA53202.1; -.
DR   EMBL; Z74823; CAA99093.1; -.
DR   PIR; S11190; PGRYI2.
DR   SGD; S0005441; IRA2.
DR   InterPro; IPR001936; RasGAP.
DR   Pfam; PF00616; RasGAP; 1.
DR   SMART; SM00323; RasGAP; 1.
DR   PROSITE; PS00509; RAS_GTPASE_ACTIV_1; 1.
DR   PROSITE; PS50018; RAS_GTPASE_ACTIV_2; 1.
KW   GTPase activation.
FT   DOMAIN     1701   1890       RAS-GAP.
FT   DOMAIN      399    409       POLY-SER.
FT   DOMAIN      412    416       POLY-SER.
FT   DOMAIN      520    528       POLY-ALA.
FT   DOMAIN     2469   2472       POLY-LEU
FT   CONFLICT   2317   2317       I -> K (IN REF. 3).
SQ   SEQUENCE   3079 AA;  351631 MW;  651EE2A2EEB47900 CRC64;


  Query Match             68.5%;  Score 37;  DB 1;  Length 3079;
  Best Local Similarity   55.6%;  Pred. No. 1.7e+02;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           || :|:| |
Db    2418 GLSYWIPNL 2426


RESULT  10
HA2Q_MOUSE
ID   HA2Q_MOUSE     STANDARD;      PRT;   221 AA.
AC   P04227;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   01-JAN-1990 (Rel. 13, Last annotation update)
DE   H-2 class II histocompatibility antigen, A-Q alpha chain (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83285339; PubMed=6309407;
RA   Benoist C.O., Mathis D.J., Kanter M.R., Williams V.E., McDevitt H.O.;
RT   "Regions of allelic hypervariability in the murine A alpha immune
RT   response gene.";
RL   Cell 34:169-177(1983).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K01925; AAA39620.1; -.
DR   PIR; A02218; HLMSA1.
DR   HSSP; P01910; 1IAK.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003597; Ig_c1
DR   InterPro; IPR001003; MHC_II_alpha.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00993; MHC_II_alpha; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   MHC II; Transmembrane; Glycoprotein.
FT   NON_TER       1      1
FT   DOMAIN       <1     76       EXTRACELLULAR ALPHA-1.
FT   DOMAIN        77    170       EXTRACELLULAR ALPHA-2.
FT   DOMAIN       171    183       CONNECTING PEPTIDE.
FT   TRANSMEM     184    209
FT   DOMAIN       210    221       CYTOPLASMIC TAIL.
FT   DISULFID      99    155       BY SIMILARITY.
FT   CARBOHYD     110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   221 AA;  24464 MW;  5C0B16EAC37D8151 CRC64;


  Query Match             66.7%;  Score 36;  DB 1;  Length 221;
  Best Local Similarity   62.5%;  Pred. No. 17;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
            |:|| ||:
Db     167 LKHWEPEI 174


RESULT  11
HA2U_MOUSE
ID   HA2U_MOUSE     STANDARD;      PRT;   227 AA.
AC   P14438;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-JAN-1990 (Rel. 13, Last annotation update)
DE   H-2 class II histocompatibility antigen, A U alpha chain (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83285339; PubMed=6309407;
RA   Benoist C.O., Mathis D.J., Kanter M.R., Williams V.E., McDevitt H.O.;
RT   "Regions of allelic hypervariability in the murine A alpha immune
RT   response gene.";
RL   Cell 34:169-177(1983).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K01926; AAA39623.1; -.
DR   HSSP; P01910; 1IAK.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003597; Ig_c1.
DR   InterPro; IPR001003; MHC_II_alpha.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00993; MHC_II_alpha; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   MHC II; Transmembrane; Glycoprotein.
FT   NON_TER       1      1
FT   DOMAIN       <1     82       EXTRACELLULAR ALPHA-1.
FT   DOMAIN        83    176       EXTRACELLULAR ALPHA 2.
FT   DOMAIN       177    189       CONNECTING PEPTIDE.
FT   TRANSMEM     190    215
FT   DOMAIN       216    227       CYTOPLASMIC TAIL.
FT   DISULFID     105    161       BY SIMILARITY.
FT   CARBOHYD     116    116       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   227 AA;  25135 MW;  D1B7EDF1B390C877 CRC64;


  Query Match             66.7%;  Score 36;  DB 1;  Length 227;
  Best Local Similarity   62.5%;  Pred. No. 18;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
            |:|| ||:
Db     173 LKHWEPEI 180


RESULT  12
HA24_HUMAN
ID   HA24_HUMAN     STANDARD;      PRT;   232 AA.
AC   P04225;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   HLA class II histocompatibility antigen, DQ(4) alpha chain precursor
DE   (DQ-DRW9 alpha chain) (Fragment).
GN   HLA-DQA1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85216512; PubMed=3879967;
RA   Moriuchi J., Moriuchi T., Silver J.;
RT   "Nucleotide sequence of an HLA-DQ alpha chain derived from a DRW9
RT   cell line. Genetic and evolutionary implications.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:3420-3424(1985).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11124; AAA59754.1; -.
DR   PIR; A02213; HLHUDQ.
DR   HSSP; P01910; 1IAK.
DR   MIM; 146880; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003597; Ig_c1.
DR   InterPro; IPR001003; MHC_II_alpha.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00993; MHC_II_alpha; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
```

```
KW   MHC II; Transmembrane; Glycoprotein; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1      4
FT   CHAIN         5    232       HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT                                DQ(4) ALPHA CHAIN.
FT   DOMAIN        5     87       EXTRACELLULAR ALPHA-1.
FT   DOMAIN       88    181       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      182    194       CONNECTING PEPTIDE
FT   TRANSMEM    195    217
FT   DOMAIN      218    232       CYTOPLASMIC TAIL.
FT   DISULFID    110    166       BY SIMILARITY.
FT   CARBOHYD     81     81       N-LINKED (GLCNAC      ) (POTENTIAL).
FT   CARBOHYD    121    121       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     84     84       A -> P (IN REF. 1; AAA59754).
FT   CONFLICT    191    191       E -> D (IN REF. 1; AAA59754).
SQ   SEQUENCE   232 AA;  25800 MW;  C51228356A711A7 CRC64;


  Query Match             66.7%;  Score 36;  DB 1;  Length 232;
  Best Local Similarity   62.5%;  Pred. No. 18;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           |:|| ||:
Db     178 LKHWEPEI 185


RESULT  13
HA2F_MOUSE
ID   HA2F_MOUSE     STANDARD;      PRT;   233 AA.
AC   P14435;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   H-2 class II histocompatibility antigen, A-F alpha chain.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83285339; PubMed=6309407;
RA   Benoist C.O., Mathis D.J., Kanter M.R., Williams V.E., McDevitt H.O.;
RT   "Regions of allelic hypervariability in the murine A alpha immune
RT   response gene.";
RL   Cell 34:169-177(1983).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K01924; -; NOT_ANNOTATED_CDS.
DR   HSSP; P01910; 1IAK.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003597; Ig_c1.
DR   InterPro; IPR001039; MHC_II_alpha.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00993; MHC_II_alpha; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   MHC II; Transmembrane; Glycoprotein; Signal.
FT   DOMAIN        1     88       EXTRACELLULAR ALPHA-1.
FT   DOMAIN       89    182       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      183    195       CONNECTING PEPTIDE.
FT   TRANSMEM    196    221
FT   DOMAIN      222    233       CYTOPLASMIC TAIL.
FT   DISULFID    111    167       BY SIMILARITY.
FT   CARBOHYD    122    122       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   233 AA;  25930 MW;  E8Db4AD44FAC79D8 CRC64;


  Query Match             66.7%;  Score 36;  DB 1;  Length 233;
  Best Local Similarity   62.5%;  Pred. No. 18;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           |:|| ||:
Db     179 LKHWEPEI 186


RESULT  14
HA2F_MOUSE
ID   HA2F_MOUSE     STANDARD;      PRT;   233 AA.
AC   P14436;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-JAN-1990 (Rel. 13, Last annotation update)
DE   H-2 class II histocompatibility antigen, A-F alpha chain.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85190610; PubMed=2581258;
RA   Landais D., Matthes H., Benoist C., Mathis D.;
RT   "A molecular basis for the Ia 2 and Ia 19 antigenic determinants.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:2930-2934(1985).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11356; AAA39621.1; -.
DR   HSSP; P01910; 1IAK.
DR   InterPro; IPR003006; Ig_MHC
DR   InterPro; IPR003597; Ig_c1.
DR   InterPro; IPR001003; MHC_II_alpha.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00993; MHC_II_alpha; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   MHC II; Transmembrane; Glycoprotein.
FT   DOMAIN        1     88       EXTRACELLULAR ALPHA-1.
FT   DOMAIN       89    182       EXTRACELLULAR ALPHA 2.
FT   DOMAIN      183    195       CONNECTING PEPTIDE.
FT   TRANSMEM    196    221
FT   DOMAIN      222    233       CYTOPLASMIC TAIL.
FT   DISULFID    111    167       BY SIMILARITY.
FT   CARBOHYD    122    122       N-LINKED (GLCNAC      ) (POTENTIAL).
SQ   SEQUENCE   233 AA;  25960 MW;  918167F9BF1b48A9 CRC64;


  Query Match             66.7%;  Score 36;  DB 1;  Length 233;
  Best Local Similarity   62.5%;  Pred. No. 18;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           |:|| ||:
Db     179 LKHWEPEI 186


RESULT  15
HA2S_MOUSE
ID   HA2S_MOUSE     STANDARD;      PRT;   233 AA.
AC   P14437;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
```

```
DT   01-JAN-1990 (Rel. 13, Last annotation update)
DE   H-2 class II histocompatibility antigen, A-S alpha chain.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85190610; PubMed 2581258;
RA   Landais D., Matthes H., Benoist C., Mathis D.;
RT   "A molecular basis for the Ia.2 and Ia.19 antigenic determinants.";
RL   Proc. Natl  Acad  Sci  U S A  82:2940-2944(1985).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11358; AAA39622.1; -.
DR   HSSP; P01910; 1IAK.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003597; Ig_c1.
DR   InterPro; IPR001003; MHC_II_alpha.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00993; MHC_II_alpha; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   MHC II; Transmembrane; Glycoprotein.
FT   DOMAIN        1     88       EXTRACELLULAR ALPHA-1.
FT   DOMAIN       89    182       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      183    195       CONNECTING PEPTIDE.
FT   TRANSMEM    196    221
FT   DOMAIN      222    233       CYTOPLASMIC TAIL.
FT   DISULFID    111    167       BY SIMILARITY.
FT   CARBOHYD    122    122       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   233 AA;  25801 MW;  6652ED01A87F3019 CRC64;


  Query Match             66.7%;  Score 36;  DB 1;  Length 233;
  Best Local Similarity   62.5%;  Pred. No. 18,
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
            |:|| ||:
Db     179 LKHWEPEI 186


Search completed: September  5, 2002, 15:31:42
Job time: 477 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:04 ; Search time 122.86 Seconds
(without alignments)
12.673 Million cell updates/sec

Title: US-09-744-804-39
Perfect score: 54
Sequence: 1 GIQHWVPEL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	54	100.0	335	4	Q9BTL9	Q9bt19 homo sapien
2	46	85.2	363	6	O77718	O77718 equus cabal
3	41	75.9	453	16	P77967	P77967 synechocyst
4	41	75.9	521	2	Q93P54	Q93p54 vibrio sp.
5	41	75.9	833	16	Q9PD09	Q9pd09 xylella fas
6	40	74.1	256	3	Q9P6Y1	Q9p6y1 neurospora
7	40	74.1	285	16	Q9KV19	Q9kv19 vibrio chol
8	40	74.1	337	12	O11317	O11317 molluscum c
9	40	74.1	452	10	Q9XIQ4	Q9xiq4 arabidopsis
10	40	74.1	480	3	O33081	O33081 mycobacteri
11	40	74.1	512	16	Q9CDD9	Q9cdd9 mycobacteri
12	40	74.1	748	12	Q98J00	Q98J00 molluscum c
13	40	74.1	1085	2	Q937Q8	Q937q8 escherichia
14	39	72.2	61	10	Q947B5	Q947b5 cucumis mel
15	39	72.2	66	5	Q9NMB6	Q9nmb6 ceratitis c
16	39	72.2	183	8	Q9XNC7	Q9xnc7 hydrodroma
17	39	72.2	192	11	Q9QWQ1	Q9qwq1 mus musculu
18	39	72.2	210	1	Q977V4	Q977v4 methanococc
19	39	72.2	388	2	Q9Z513	Q9z513 streptomyce
20	39	72.2	357	16	Q9HY94	Q9hy94 pseudomonas
21	39	72.2	426	11	Q9WTS3	Q9wts3 mus musculu
22	39	72.2	463	11	Q9R1X9	Q9r1x9 mus musculu
23	39	72.2	481	16	Q9HVD2	Q9hvd2 pseudomonas
24	39	72.2	1155	16	Q9A782	Q9a782 caulobacter
25	38	70.4	129	16	Q99QY1	Q99qy1 staphylococ
26	38	70.4	183	8	Q9XND2	Q9xnd2 erythracaru
27	38	70.4	238	16	Q9A510	Q9a510 caulobacter
28	38	70.4	288	16	Q9CP85	Q9cp85 pasteurella
29	38	70.4	635	10	Q91DJ5	Q91dj5 lycopersico
30	38	70.4	665	5	Q9W241	Q9w241 drosophila
31	38	70.4	740	3	O94154	O94154 pleurotus o
32	38	70.4	1335	4	Q9P222	Q9p222 homo sapien
33	38	70.4	4025	4	Q9NR13	Q9nr13 homo sapien
34	37	68.5	67	10	Q42670	Q42670 arabidopsis
35	37	68.5	112	10	Q9FWL5	Q9fwl5 oryza sat.7
36	37	68.5	122	8	O22168	O22168 chlorella v
37	37	68.5	134	7	Q95548	Q95548 sus scrofa
38	37	68.5	158	13	O93475	O93475 xenopus lae
39	37	68.5	158	13	O93476	O93476 xenopus lae
40	37	68.5	183	8	Q9XNC8	Q9xnc8 halacaropsi
41	37	68.5	223	4	Q9H0G3	Q9h0g3 homo sapien
42	37	68.5	260	12	Q67685	Q67685 groundnut r
43	37	68.5	274	4	Q96M21	Q96m21 homo sapien
44	37	68.5	286	2	Q46954	Q46954 escherichia
45	37	68.5	286	2	O34176	O34176 klebsiella

ALIGNMENTS

RESULT 1
Q9BTL9
ID Q9BTL9 PRELIMINARY; PRT; 335 AA.
AC Q9BTL9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO MILK FAT GLOBULE-EGF FACTOR 8 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA.;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
DR EMBL; BC003610; AAH03610.1; -.
DR HSSP; P08709; 1BF9.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR000421; FA58_C.
DR PRINTS; PR00010; EGFBLOOD.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00231; EGF_like; 1.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01285; FA58C_1; 1.
DR PROSITE; PS01286; FA58C_2; 2.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 335 AA, 37523 MW, 20584E5DB78EBBE25 CRC64;

Query Match 100.0%; Score 54; DB 4; Length 335;
Best Local Similarity 100.0%; Pred. No. 9.1e;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GLQHWVPEL 9
           |||||||||
Db      97 GLQHWVPEL 105


RESULT   2
O77718
ID   O77718      PRELIMINARY;      PRT;   363 AA.
AC   O77718;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SPERM-MEMBRANE ASSOCIATED PROTEIN P47 (FRAGMENT).
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Gentzel M., Toepfer-Petersen E.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ010121; CAA09010.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 1.
KW   EGF-like domain; Glycoprotein.
FT   NON_TER       1       1
FT   NON_TER     363     363
SQ   SEQUENCE   363 AA;  40744 MW;  1F8B6395AF323380 CRC64;


  Query Match             85.2%;  Score 46;  DB 6;  Length 363;
  Best Local Similarity   88.9%;  Pred. No. 3.1;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           ||| |||||
Db     102 GLQRWVPEL 110


RESULT   3
P77967
ID   P77967      PRELIMINARY;      PRT;   453 AA.
AC   P77967;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   DNA PHOTOLYASE.
GN   PHR OR SLL1629.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D90909; BAA17766.1; -.
DR   HSSP; P05327; 1QNF.
DR   InterPro; IPR000474; DNA_photolyase.
DR   InterPro; IPR002081; DNA_photolyase_1.
DR   Pfam; PF00875; DNA_photolyase; 1.
DR   ProDom; PD004390; DNA_photolyase; 1.
DR   PROSITE; PS00394; DNA_PHOTOLYASES_1_1; 1.
KW   Lyase; Complete proteome.
SQ   SEQUENCE   453 AA;  52920 MW;  C8E55B60C025E5B2 CRC64;


  Query Match             75.9%;  Score 41;  DB 16;  Length 453;
  Best Local Similarity   75.0%;  Pred. No. 29;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           |:|||:|||
Db     392 LPHWLPEL 399


RESULT   4
Q93P54
ID   Q93P54      PRELIMINARY;      PRT;   521 AA.
AC   Q93P54;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ALKALINE PHOSPHATASE.
OS   Vibrio sp. G15-21.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=169049;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=G15-21;
RA   Asgeirsson B., Andresson O.S.;
RT   "Primary structure of cold-adapted alkaline phosphatase from a Vibrio
RT   sp., as deduced from the nucleotide sequence.";
RL   Biochim. Biophys. Acta 1550:8(2001).
DR   EMBL; AF352014; AAK94204.1; -.
SQ   SEQUENCE   521 AA;  57384 MW;  9D0493DB4D0CF192 CRC64;


  Query Match             75.9%;  Score 41;  DB 2;  Length 521;
  Best Local Similarity   62.5%;  Pred. No. 33;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPE 8
           ||:||:|:
Db     170 GLRHWIPK 177


RESULT   5
Q9PDZ9
ID   Q9PDZ9      PRELIMINARY;      PRT;   833 AA.
AC   Q9PDZ9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ATP-DEPENDENT HELICASE.
GN   XF1229.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
```

```
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
CC   -!- SIMILARITY: TO HELICASE C-TERMINAL DOMAIN
DR   EMBL; AE003956; AAF84039.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Complete proteome; Helicase.
SQ   SEQUENCE   833 AA;  92557 MW;  BCCAB0684C082C8A CRC64;

  Query Match             75.9%;  Score 41;  DB 16;  Length 833;
  Best Local Similarity   75.0%;  Pred. No. 53;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LQHWVPEL 9
              |:||:|||
Db        658 LRHWMPEL 665

RESULT 6
Q9P6Y1
ID   Q9P6Y1      PRELIMINARY;      PRT;   256 AA.
AC   Q9P6Y1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   13E11.350.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL353820; CAB88603.1; -.
DR   InterPro; IPR001601; Meth-transf.
KW   Hypothetical protein.
SQ   SEQUENCE   256 AA;  27789 MW;  5251FBC58B6BDDD9 CRC64;

  Query Match             74.1%;  Score 40;  DB 3;  Length 256;
  Best Local Similarity   66.7%;  Pred. No. 24;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GLQHWVPEL 9
               ||: ||||:
Db        141 GLEEWVPEV 149

RESULT 7
Q9KV19
ID   Q9KV19      PRELIMINARY;      PRT;   285 AA.
AC   Q9KV19;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PHOSPHATIDYLSERINE DECARBOXYLASE.
GN   VC0339.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004122; AAF93512.1; -.
DR   TIGR; VC0339; -.
DR   InterPro; IPR003817; PS_Dcarbxylase.
DR   Pfam; PF02666; PS_Dcarbxylase; 1.
KW   Complete proteome.
SQ   SEQUENCE   285 AA;  31651 MW;  9BB3A4E8EAE7383B CRC64;

  Query Match             74.1%;  Score 40;  DB 16;  Length 285;
  Best Local Similarity   62.5%;  Pred. No. 27;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GLQHWVPE 8
              |||:|:|:
Db          7 GLQYWIPQ 14

RESULT 8
O11317
ID   O11317      PRELIMINARY;      PRT;   337 AA.
AC   O11317;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   H1-60.1 PROTEIN (FRAGMENT).
GN   H1-60.1.
OS   Molluscum contagiosum virus subtype 1 (MCVI).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Molluscipoxvirus.
OX   NCBI_TaxID=10280;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Moratilla M., Agromayor M., Nunez A., Funes J.M., Varas A.J.,
RA   Lopez Estebaranz J.L., Esteban M., Martin-Gallardo A.;
RT   "A Random DNA Sequencing, Computer Based Approach for the Generation
RT   of a Gene Map of Molluscum Contagiosum Virus.";
RL   Virus Genes 0:0-0(1997).
DR   EMBL; U86900; AAB57947.1; -.
FT   NON_TER     337    337
SQ   SEQUENCE   337 AA;  38567 MW;  DC28C3D828785A7BE CRC64;
```

```
  Query Match             74.1%;  Score 40;  DB 12;  Length 337;
  Best Local Similarity   85.7%;  Pred. No. 32;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 QHWVPEL 9
           :||||||
Db      52 RHWVPEL 58


RESULT   9
Q9XIQ4
ID   Q9XIQ4      PRELIMINARY;      PRT;   452 AA.
AC   Q9XIQ4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIMILAR TO FLAVONOL 3-O-GLUCOSYLTRANSFERASE.
GN   F13O11.22.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA   Altafi H., Araujo R., Bolivar L., Bowley D., Buehler E., Dunn P.,
RA   Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA   Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA   Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006193; AAD38266.1; -.
KW   Transferase.
SQ   SEQUENCE   452 AA;  50554 MW;  9E7EB6C02111E413 CRC64;


  Query Match             74.1%;  Score 40;  DB 10;  Length 452;
  Best Local Similarity   75.0%;  Pred. No. 43;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           | ||||||
Db     115 LAHWVPEM 122


RESULT  10
O33081
ID   O33081      PRELIMINARY;      PRT;   480 AA.
AC   O33081;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 51.4 KDA PROTEIN.
GN   MLCB628.10C.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Eiglmeier K., Garnier T., De Rossi E., Fsihi H., Cole S.T.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93188700; PubMed=8446027;
RA   Eiglmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT   "Use of an ordered cosmid library to deduce the genomic organization
RT   of Mycobacterium leprae.";
RL   Mol. Microbiol. 7:197-206(1993).
DR   EMBL; Y14967; CAA75198.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   480 AA;  51400 MW;  34B19BD4567F55F CRC64;


  Query Match             74.1%;  Score 40;  DB 2;  Length 480;
  Best Local Similarity   85.7%;  Pred. No. 45;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GLQHWVP 7
           | |||||
Db     247 GYQHWVP 253


RESULT  11
Q9CDD9
ID   Q9CDD9      PRELIMINARY;      PRT;   512 AA.
AC   Q9CDD9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   PUTATIVE MEMBRANE PROTEIN.
GN   ML0047.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
DR   EMBL; AL583917; CAC29555.1; -.
DR   Leproma; ML0047; -.
KW   Complete proteome.
SQ   SEQUENCE   512 AA;  54503 MW;  FF7435DD39D6A800 CRC64;


  Query Match             74.1%;  Score 40;  DB 16;  Length 512;
  Best Local Similarity   85.7%;  Pred. No. 48;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GLQHWVP 7
           | |||||
Db     279 GYQHWVP 285


RESULT  12
Q98200
ID   Q98200      PRELIMINARY;      PRT;   748 AA.
AC   Q98200;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MC032L.
GN   MC032L.
OS   Molluscum contagiosum virus subtype 1 (MCVI).
OC   Viruses, dsDNA viruses, no RNA stage, Poxviridae, Chordopoxvirinae;
OC   Molluscipoxvirus.
OX   NCBI_TaxID=10280;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96325459; PubMed=8670425;
RA   Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,
RA   Moss B.;
RT   "Genome sequence of a human tumorigenic poxvirus: Prediction of
```

```
RT   specific host response-evasion genes.";
RL   Science 273:813-816(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.;
RA   Moss B.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U60315; AAC55160.1; -.
SQ   SEQUENCE   748 AA;  84411 MW;  61A2741F62C0A5BA CRC64;


  Query Match             74.1%;  Score 40;  DB 12;  Length 748;
  Best Local Similarity   85.7%;  Pred. No. 71;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 QHWVPEL 9
           :||||||
Db      52 EHWVPEL 58


RESULT  13
Q937Q8
ID   Q937Q8      PRELIMINARY;      PRT;  1005 AA.
AC   Q937Q8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE TRANSPOSASE.
OS   Escherichia coli.
OG   Plasmid pMSP071
OC   Bacteria; Proteobacteria, gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=785-D; TRANSPOSON=IN60;
RX   MEDLINE=20316819; PubMed=10858363;
RA   Sabate M., Tarrago R., Navarro F., Miro E., Verges C., Barbe J.,
RA   Prats G.;
RT   "Cloning and sequence of the gene encoding a novel cefotaximase
RT   hydrolysing beta-lactamase (CTX-M-9) from Escherichia coli in Spain.";
RL   Antimicrob. Agents Chemother. 44:1970-1973(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=785-D; TRANSPOSON=IN60;
RA   Sabate M., Navarro F., Barbe J., Miro E., Mirelis B., Prats G.;
RT   "A novel complex sul1 type integron (In60) in Escherichia coli
RT   carrying the blaCTX-M-9 ";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF174129; AAK60193.1; -.
KW   Plasmid.
SQ   SEQUENCE   1005 AA;  116157 MW;  726E107EFADA057F CRC64;


  Query Match             74.1%;  Score 40;  DB 2;  Length 1005;
  Best Local Similarity   62.5%;  Pred. No. 96;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           :|||:|||:
Db     234 IQHWMPEV 241


RESULT  14
Q947B5
ID   Q947B5      PRELIMINARY;      PRT;    61 AA.
AC   Q947B5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CRYPTOCHROME 2 (FRAGMENT).
GN   CRY2.
OS   Cucumis melo (Muskmelon).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta, Magnoliophyta, eudicotyledons; core eudicots; Rosidae;
OC   eurosids I, Cucurbitales, Cucurbitaceae; Cucumis.
OX   NCBI_TaxID=3656;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Perrotta G., Salamini F., Nebulose E., Bozzi L., Giuliano G.;
RT   "Tomato and barley contain duplicated copies of cryptochrome 1.";
RL   Plant Cell Environ. 24:991-997(2001).
DR   EMBL; AF348457; AAL02088.1; -.
FT   NON_TER       1       1
FT   NON_TER      61      61
SQ   SEQUENCE   61 AA;  7012 MW;  1432B3B23C69DC9D CRC64;


  Query Match             72.2%;  Score 39;  DB 10;  Length 61;
  Best Local Similarity   62.5%;  Pred. No. 8.3;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           ::||:|||
Db      48 IRHWLPEL 55


RESULT  15
Q9NMB6
ID   Q9NMB6      PRELIMINARY;      PRT;    66 AA.
AC   Q9NMB6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PROBABLE R10H10.1 PROTEIN (FRAGMENT).
GN   LM26.176.
OS   Leishmania major.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5664;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FRIEDLIN;
RA   Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL160493; CAB97794.1; -.
FT   NON_TER      66      66
SQ   SEQUENCE   66 AA;  7396 MW;  7EC60CDBA352AD7C CRC64;


  Query Match             72.2%;  Score 39;  DB 5;  Length 66;
  Best Local Similarity   62.5%;  Pred. No. 9;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           | |||:||:
Db      26 LMHWIPEV 33



Search completed: September 5, 2002, 15:31:06
Job time: 471 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:27:35 ; Search Time 58.86 Seconds
(without alignments)
3.735 Million cell updates/sec

Title: US-09-744-804-39
Perfect score: 54
Sequence: 1 GLQHWVPEL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	54	100.0	160	2	US-08-162-402B-10	Sequence 10, Appl
2	54	100.0	487	2	US-08-162-402B-6	Sequence 6, Appli
3	54	100.0	495	2	US-08-162-402B-8	Sequence 8, Appli
4	39	72.2	160	2	US-08-162-402B-11	Sequence 11, Appl
5	39	72.2	320	2	US-08-480-229C-20	Sequence 20, Appl
6	39	72.2	320	2	US-08-659-235C-20	Sequence 20, Appl
7	39	72.2	463	2	US-08-162-402B-9	Sequence 9, Appli
8	37	68.5	264	2	US-08-710-697-8	Sequence 8, Appli
9	37	68.5	264	4	US-08-727-616A-8	Sequence 8, Appli
10	37	68.5	265	2	US-08-710-697-6	Sequence 6, Appli
11	37	68.5	265	4	US-08-727-616A-6	Sequence 6, Appli
12	37	68.5	3079	5	PCT-US94-00198-4	Sequence 4, Appli
13	36	66.7	255	2	US-08-484-905-110	Sequence 110, App
14	36	66.7	255	3	US-08-481-985B-110	Sequence 110, App
15	36	66.7	255	4	US-08-370-476-110	Sequence 110, App
16	36	66.7	256	2	US-08-484-905-113	Sequence 113, App
17	36	66.7	256	3	US-08-481-985B-113	Sequence 113, App
18	36	66.7	256	4	US-08-370-476-113	Sequence 113, App
19	36	66.7	500	4	US-08-960-190A-25	Sequence 25, Appl
20	36	66.7	514	1	US-07-872-644-17	Sequence 17, Appl
21	36	66.7	514	1	US-08-297-494-17	Sequence 17, Appl
22	36	66.7	514	1	US-08-297-510-17	Sequence 17, Appl
23	36	66.7	514	1	US-08-479-532-17	Sequence 17, Appl
24	36	66.7	514	1	US-08-455-526-17	Sequence 17, Appl
25	36	66.7	514	1	US-08-455-525-17	Sequence 17, Appl
26	36	66.7	514	3	US-09-139-491-17	Sequence 17, Appl
27	36	66.7	514	5	PCT-US92-03222-17	Sequence 17, Appl
28	36	66.7	530	1	US-07-872-644-6	Sequence 6, Appli
29	36	66.7	530	1	US-08-297-494-6	Sequence 6, Appli
30	36	66.7	530	1	US-08-297-510-6	Sequence 6, Appli
31	36	66.7	530	1	US-08-479-532-6	Sequence 6, Appli
32	36	66.7	530	1	US-08-455-526-6	Sequence 6, Appli
33	36	66.7	530	1	US-08-455-525-6	Sequence 6, Appli
34	36	66.7	530	3	US-09-139-491-6	Sequence 6, Appli
35	36	66.7	530	5	PCT-US92-03222-6	Sequence 6, Appli
36	36	66.7	535	1	US-07-872-644-49	Sequence 49, Appl
37	36	66.7	535	1	US-08-297-494-49	Sequence 49, Appl
38	36	66.7	535	1	US-08-297-510-49	Sequence 49, Appl
39	36	66.7	535	1	US-08-479-532-49	Sequence 49, Appl
40	36	66.7	535	1	US-08-455-526-49	Sequence 49, Appl
41	36	66.7	535	1	US-08-455-525-49	Sequence 49, Appl
42	36	66.7	535	3	US-09-139-491-49	Sequence 49, Appl
43	36	66.7	535	5	PCT-US92-03222-49	Sequence 49, Appl
44	36	66.7	674	1	US-08-803-972-7	Sequence 7, Appli
45	36	66.7	674	1	US-08-803-972-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-162-402B-10
; Sequence 10, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 29,230
; REFERENCE/DOCKET NUMBER: PBC 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 160 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-10

Query Match 100.0%; Score 54; DB 2; Length 160;
Best Local Similarity 100.0%; Pred. No. 0.015;

```
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           |||||||||
Db      29 GLQHWVPEL 37


RESULT   2
US-08-162-402B-6
; Sequence 6, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  387 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-6


  Query Match             100.0%;  Score 54;  DB 2;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.039;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           |||||||||
Db      97 GLQHWVPEL 105


RESULT   3
US-08-162-402B-8
; Sequence 8, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213 622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  465 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-8


  Query Match             100.0%;  Score 54;  DB 2;  Length 465;
  Best Local Similarity   100.0%;  Pred. No. 0.047;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           |||||||||
Db     175 GLQHWVPEL 183


RESULT   4
US-08-162-402B-11
; Sequence 11, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
```

```
;       FILING DATE:  03-DEC-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Amzel, Viviana
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  P66 38215
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  213-622-7700
;       TELEFAX:  213-489-4210
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  160 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
US-08-162-402B-11


  Query Match             72.2%;  Score 39;  DB 2;  Length 160;
  Best Local Similarity   77.8%;  Pred. No. 6.7;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           ||| | |||
Db      29 GLQRWGPEL 37


RESULT   5
US-08-480-229C-20
; Sequence 20, Application US/08480229C
; Patent No. 5874562
;  GENERAL INFORMATION:
;     APPLICANT:  Quertermous, Thomas
;     APPLICANT:  Hogan, Brigid
;     APPLICANT:  Snodgrass, H. Ralph
;     APPLICANT:  Zupancic, Thomas J.
;     TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;     TITLE OF INVENTION:  CELL LOCUS 1
;     NUMBER OF SEQUENCES:  29
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds LLP
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  United States
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/480,229C
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Poissant, Brian M.
;       REGISTRATION NUMBER:  28,462
;       REFERENCE/DOCKET NUMBER:  8907-0026-999
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-8864/9741
;       TELEX:  66141 Pennie
;   INFORMATION FOR SEQ ID NO:  20:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  320 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  protein
US-08-480-229C-20


  Query Match             72.2%;  Score 39;  DB 2;  Length 320;
  Best Local Similarity   77.8%;  Pred. No. 14;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           ||| | |||
Db      28 GLQRWGPEL 36


RESULT   6
US-08-659-235C-20
; Sequence 20, Application US/08659235C
; Patent No. 5877281
;  GENERAL INFORMATION:
;     APPLICANT:  Quertermous, Thomas
;     APPLICANT:  Hogan, Brigid
;     APPLICANT:  Snodgrass, H. Ralph
;     APPLICANT:  Zupancic, Thomas J.
;     TITLE OF INVENTION:  DEVELOPMENTALLY REGULATED ENDOTHELIAL
;     TITLE OF INVENTION:  CELL LOCUS-1
;     NUMBER OF SEQUENCES:  29
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds LLP
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  United States
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/659,235C
;       FILING DATE:  05-JUN-1996
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Poissant, Brian M.
;       REGISTRATION NUMBER:  28,462
;       REFERENCE/DOCKET NUMBER:  8907-0034-999
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-8864/9741
;       TELEX:  66141 Pennie
;   INFORMATION FOR SEQ ID NO:  20:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  320 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  protein
US-08-659-235C-20


  Query Match             72.2%;  Score 39;  DB 2;  Length 320;
  Best Local Similarity   77.8%;  Pred. No. 14;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           ||| | |||
Db      28 GLQRWGPEL 36


RESULT   7
US-08-162-402B-9
; Sequence 9, Application US/08162402B
; Patent No. 5972337
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  IAROCCA, DAVID J
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  463 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-9


  Query Match             72.2%;  Score 39;  DB 2;  Length 463;
  Best Local Similarity   77.8%;  Pred. No. 21;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
            ||| | |||
Db     175 GLQRWGPEL 183


RESULT   8
US-08-719-697-8
; Sequence 8, Application US/08719697
; Patent No. 5928888
;  GENERAL INFORMATION:
;    APPLICANT:  Whitney, Michael A.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SENSITIVE
;    TITLE OF INVENTION:  AND RAPID, FUNCTIONAL IDENTIFICATION OF GENOMIC
;    TITLE OF INVENTION:  POLYNUCLEOTIDES AND SECONDARY SCREENING CAPABILITIES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,697
;      FILING DATE:  26-SEP-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  08366/003001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  264 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-719-697-8


  Query Match             68.5%;  Score 37;  DB 2;  Length 264;
  Best Local Similarity   75.0%;  Pred. No. 26;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
             | || |||
Db     138 LDHWEPEL 145


RESULT   9
US-08-727-616A-8
; Sequence 8, Application US/08727616A
; Patent No. 6291162
;  GENERAL INFORMATION:
;    APPLICANT:  Tsien, Roger Y.
;    APPLICANT:  Zlokarnik, Gregor
;    TITLE OF INVENTION:  SUBSTRATES FOR BETA-LACTAMASE
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/727,616A
;      FILING DATE:  15-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/407,544
;      FILING DATE:  20-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07257/034001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  264 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
US-08-727-616A-8
```

```
  Query Match             68.5%;  Score 37;  DB 4;  Length 264;
  Best Local Similarity   75.0%;  Pred. No. 26;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           | || |||
Db     138 LDHWEPEL 145


RESULT  10
US-08-719-697-6
; Sequence 6, Application US/08719697
; Patent No. 5928888
;  GENERAL INFORMATION:
;    APPLICANT:  Whitney, Michael A.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SENSITIVE
;    TITLE OF INVENTION:  AND RAPID, FUNCTIONAL IDENTIFICATION OF GENOMIC
;    TITLE OF INVENTION:  POLYNUCLEOTIDES AND SECONDARY SCREENING CAPABILITIES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,697
;      FILING DATE:  26-SEP-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07257/003001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  265 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-719-697-6


  Query Match             68.5%;  Score 37;  DB 2;  Length 265;
  Best Local Similarity   75.0%;  Pred. No. 26;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           | || |||
Db     139 LDHWEPEL 146


RESULT  11
US-08-727-616A-6
; Sequence 6, Application US/08727616A
; Patent No. 6291162
;  GENERAL INFORMATION:
;    APPLICANT:  Tsien, Roger Y.
;    APPLICANT:  Zlokarnik, Gregor
;    TITLE OF INVENTION:  SUBSTRATES FOR BETA-LACTAMASE
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/727,616A
;      FILING DATE:  15-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/407,544
;      FILING DATE:  20-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07257/034001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  265 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
US-08-727-616A-6


  Query Match             68.5%;  Score 37;  DB 4;  Length 265;
  Best Local Similarity   75.0%;  Pred. No. 26;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           | || |||
Db     139 LDHWEPEL 146


RESULT  12
PCT-US94-00198-4
; Sequence 4, Application PC/TUS9400198
;  GENERAL INFORMATION:
;    APPLICANT:  Schering Corp.
;    TITLE OF INVENTION:  RAS Associated GAP Proteins
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  1 Giralda Farms
;      CITY:  Madison
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  94304-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  6.0.8
;      SOFTWARE:  Microsoft Word 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/00198
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/004,824
;      FILING DATE:  15-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lunn, Paul G.
;      REGISTRATION NUMBER:  32,743
;      REFERENCE/DOCKET NUMBER:  EX0353 PCT
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  (201)822-7255
;       TELEFAX:  (201)822-7039
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3079 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     ORIGINAL SOURCE:
;       ORGANISM:  Saccharomyces cerevisiae
PCT-US94-00198-4


  Query Match             68.5%;  Score 37,  DB 5,  Length 3079,
  Best Local Similarity   55.6%;  Pred. No. 3.9e+02;
  Matches    5;  Conservative    2,  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GLQHWVPEL 9
           || :|:| |
Db    2418 GLSYWIPNL 2426


RESULT  14
US-08-484-905-110
; Sequence 110, Application US/08484905
; Patent No. 5976551
;  GENERAL INFORMATION:
;    APPLICANT:  Mottez, Estelle
;    APPLICANT:  Abastado, Jean-Pierre
;    APPLICANT:  Kourilsky, Philippe
;    TITLE OF INVENTION:  An Altered Major Histocompatibility
;    TITLE OF INVENTION:  Complex(MHC) Determinant and Methods for Using the
;    TITLE OF INVENTION:  Determinant
;    NUMBER OF SEQUENCES:  127
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS-/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,905
;      FILING DATE:  07-JUNE-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/801,818
;      FILING DATE:  05-DEC-1991
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/792,473
;      FILING DATE:  15-NOV-1991
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Potter, Jane E. R.
;      REGISTRATION NUMBER:  33,332
;      REFERENCE/DOCKET NUMBER:  03495.0106-03000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  110:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  255 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-484-905-110


  Query Match             66.7%,  Score 36,  DB 2;  Length 255;
  Best Local Similarity   62.5%;  Pred. No. 38;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           |:|| ||:
Db     201 LKHWEPEI 208


RESULT  14
US-08-481-985B-110
; Sequence 110, Application US/08481985B
; Patent No. 6011146
;  GENERAL INFORMATION:
;    APPLICANT:  Mottez, Estelle
;    APPLICANT:  Abastado, Jean-Pierre
;    APPLICANT:  Kourilsky, Phillipe
;    TITLE OF INVENTION:  Altered Major Histocompatibility Complex
;    TITLE OF INVENTION:
;    NUMBER OF SEQUENCES:  148
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/481,985B
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/801,818
;      FILING DATE:  05-DEC-1991
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/792,473
;      FILING DATE:  15-NOV-1991
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0106-04000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  110:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  255 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-481-985B-110


  Query Match             66.7%,  Score 36,  DB 3;  Length 255;
  Best Local Similarity   62.5%,  Pred. No. 38;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
           |:|| ||:
Db     201 LKHWEPEI 208


RESULT  15
US-08-370-476-110
```

```
; Sequence 110, Application US/08370476
; Patent No. 6153408
;  GENERAL INFORMATION:
;     APPLICANT:  Mottez, Estelle
;     APPLICANT:  Abastado, Jean-Pierre
;     APPLICANT:  Kourilsky, Phillipe
;     APPLICANT:  Lone, Yu-Chun
;     APPLICANT:  Ojcius, David
;     APPLICANT:  Casrouge, Armanda
;     TITLE OF INVENTION:  Altered Major Histocompatibility Complex
;     TITLE OF INVENTION:
;     NUMBER OF SEQUENCES:  127
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;       ADDRESSEE:  Dunner
;       STREET:  1300 I Street, N.W., Suite 700
;       CITY:  Washington
;       STATE:  D.C.
;       ZIP:  20005-3315
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/370,476
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/117,575
;       FILING DATE:  07-SEP-1993
;       APPLICATION NUMBER:  US 08/072,787
;       FILING DATE:  06-JUN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/801,818
;       FILING DATE:  05-DEC-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/792,473
;       FILING DATE:  15-NOV-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meyers, Kenneth J.
;       REGISTRATION NUMBER:  25,146
;       REFERENCE/DOCKET NUMBER:  0524 0001-01000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-408-4000
;       TELEFAX:  202-408-4400
;   INFORMATION FOR SEQ ID NO:  110:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  255 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-370-476-110


  Query Match             66.7%;  Score 36;  DB 4;  Length 255;
  Best Local Similarity   62.5%;  Pred. No. 38;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LQHWVPEL 9
            |:|| ||:
Db     201 LKHWEPEI 208



Search completed: September  5, 2002, 15:27:36
Job time: 321 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:27:46 ; Search time 59.86 Seconds
(without alignments)
3.735 Million cell updates/sec

Title: US-09-744-804-40
Perfect score: 43
Sequence: 1 VQFVASYKV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfile1.pep:*

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	100.0	85	2	US-08-480-229C-2	Sequence 2, Appli
2	43	100.0	85	2	US-08-659-235C-2	Sequence 2, Appli
3	43	100.0	159	2	US-08-162-402B-12	Sequence 12, Appl
4	43	100.0	217	1	US-07-607-538C-3	Sequence 3, Appli
5	43	100.0	217	2	US-08-162-402B-3	Sequence 3, Appli
6	43	100.0	218	1	US-07-607-538C-2	Sequence 2, Appli
7	43	100.0	218	2	US-08-162-402B-2	Sequence 2, Appli
8	43	100.0	387	2	US-08-162-402B-6	Sequence 6, Appli
9	43	100.0	455	2	US-08-162-402B-8	Sequence 8, Appli
10	36	83.7	321	2	US-08-480-229C-21	Sequence 21, Appl
11	36	83.7	321	2	US-08-659-235C-21	Sequence 21, Appl
12	36	83.7	480	2	US-08-480-229C-10	Sequence 10, Appl
13	36	83.7	480	2	US-08-659-235C-10	Sequence 10, Appl
14	36	83.7	513	2	US-08-480-229C-14	Sequence 14, Appl
15	36	83.7	513	2	US-08-659-235C-14	Sequence 14, Appl
16	34	79.1	109	1	US-08-111-939-25	Sequence 25, Appl
17	34	79.1	157	2	US-08-162-402B-13	Sequence 13, Appl
18	34	79.1	320	2	US-08-480-229C-20	Sequence 20, Appl
19	34	79.1	320	2	US-08-659-235C-20	Sequence 20, Appl
20	34	79.1	463	2	US-08-162-402B-9	Sequence 9, Appli
21	29	67.4	85	2	US-08-480-229C-1	Sequence 1, Appli
22	29	67.4	85	2	US-08-480-229C-5	Sequence 5, Appli
23	29	67.4	85	2	US-08-480-229C-7	Sequence 7, Appli
24	29	67.4	85	2	US-08-480-229C-8	Sequence 8, Appli
25	29	67.4	85	2	US-08-659-235C-1	Sequence 1, Appli
26	29	67.4	85	2	US-08-659-235C-5	Sequence 5, Appli
27	29	67.4	85	2	US-08-659-235C-7	Sequence 7, Appli
28	29	67.4	85	2	US-08-659-235C-8	Sequence 8, Appli
29	29	67.4	104	1	US-08-111-939-26	Sequence 26, Appl
30	29	67.4	159	2	US-08-162-402B-27	Sequence 27, Appl
31	29	67.4	156	2	US-08-162-402B-18	Sequence 18, Appl
32	29	67.4	181	5	PCT-US96-01314-61	Sequence 61, Appl
33	29	67.4	184	3	US-08-974-899-7	Sequence 7, Appli
34	29	67.4	184	3	US-08-974-899-8	Sequence 8, Appli
35	29	67.4	1065	4	US-08-640-172-9	Sequence 9, Appli
36	29	67.4	1065	4	US-09-375-419-9	Sequence 9, Appli
37	29	67.4	1170	2	US-08-789-078-2	Sequence 2, Appli
38	29	67.4	1170	2	US-08-752-633-2	Sequence 2, Appli
39	29	67.4	1170	2	US-08-476-062A-42	Sequence 42, Appl
40	29	67.4	1170	5	PCT-US95-04886-2	Sequence 2, Appli
41	29	67.4	1170	5	PCT-US96-01314-42	Sequence 42, Appl
42	29	67.4	1477	3	US-08-492-459-10	Sequence 10, Appl
43	29	67.4	1477	3	US-08-423-752-10	Sequence 10, Appl
44	29	67.4	1477	4	US-08-945-994-3	Sequence 3, Appli
45	29	67.4	1477	4	US-08-710-874-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-480-229C-2
; Sequence 2, Application US/08480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Sweetgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,229C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: [illegible]
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-480-229C-2

Query Match 100.0%, Score 43, DB 2, Length 85,
Best Local Similarity 100.0%, Pred. No. 0.83,
Matches 9, Conservative 0, Mismatches 0, Indels 0, Gaps 0,

```
Qy       1 VQFVASYKV 9
           |||||||||
Db      24 VQFVASYKV 32


RESULT   2
US-08-659-235C-2
; Sequence 2, Application US/08659235C
; Patent No. 5877281
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS 1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,235C
;      FILING DATE:  05-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8007-6034-606
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  85 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-659-235C-2


  Query Match             100.0%;  Score 43;  DB 2;  Length 85;
  Best Local Similarity   100.0%;  Pred. No. 0.03;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||||||||
Db      24 VQFVASYKV 32


RESULT   3
US-08-162-402B-12
; Sequence 12, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St , 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  159 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-12


  Query Match             100.0%;  Score 43;  DB 2;  Length 159;
  Best Local Similarity   100.0%;  Pred. No. 0.058;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||||||||
Db      85 VQFVASYKV 93


RESULT   4
US-07-607-538C-3
; Sequence 3, Application US/07607538C
; Patent No. 5455031
;  GENERAL INFORMATION:
;    APPLICANT:  Ceriani Dr., Roberto L.
;    APPLICANT:  Peterson Dr., Jerry A.
;    APPLICANT:  Larocca, David J.
;    TITLE OF INVENTION:  POLYPEPTIDE WITH 46
;    TITLE OF INVENTION:  DIFFERENTIATION ANTIGEN BINDING SPECIFITY AND  CLOTTING
;    TITLE OF INVENTION:  FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
;    TITLE OF INVENTION:  FUSION  PROTEIN,  POLYNUCLEOTIDE AND  POLYRIBO-
;    TITLE OF INVENTION:  NUCLEOTIDE  ENCODING  THE POLYPEPTIDE, ANTI-
;    TITLE OF INVENTION:  POLYPEPTIDE ANTIBODIES, KITS  AND  METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  V.  Amzel & Assoc.
;      STREET:  2255 No. 54559311th Broadway
;      CITY:  Walnut Creek
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94596
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS 5.0
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/07/607,538C
;       FILING DATE:  01-NOV-1990
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Viviana Amzel
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  CRFCC-004
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (510) 943-1931
;       TELEFAX:  (510) 943-1189
;       TELEX:  N.A.
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  217 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     FRAGMENT TYPE:
US-07-607-538C-3


  Query Match           100.0%;  Score 43;  DB 1;  Length 217;
  Best Local Similarity  100.0%;  Pred. No. 0.081;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||||||||
Db     143 VQFVASYKV 151


RESULT   5
US-08-162-402B-3
; Sequence 3, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;     APPLICANT:  CERIANI, ROBERTO L.
;     APPLICANT:  PETERSON, JERRY A.
;     APPLICANT:  LAROCCA, DAVID J.
;     TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;     TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;     NUMBER OF SEQUENCES:  29
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pretty, Schroeder & Poplawski
;       STREET:  444 South Flower St., 19th Floor
;       CITY:  Los Angeles
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/162,402B
;       FILING DATE:  03-DEC-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Amzel, Viviana
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  PD0 38215
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  213-622-7700
;       TELEFAX:  213-489-4210
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  217 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
US-08-162-402B-3


  Query Match           100.0%;  Score 43;  DB 2;  Length 217;
  Best Local Similarity  100.0%;  Pred. No. 0.081;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||||||||
Db     143 VQFVASYKV 151


RESULT   6
US-07-607-538C-2
; Sequence 2, Application US/07607538C
; Patent No. 5455031
;  GENERAL INFORMATION:
;     APPLICANT:  Ceriani Dr., Roberto L.
;     APPLICANT:  Peterson Dr., Jerry A.
;     APPLICANT:  Larocca, David J.
;     TITLE OF INVENTION:  POLYPEPTIDE WITH 46
;     TITLE OF INVENTION:  DIFFERENTIATION ANTIGEN BINDING SPECIFITY AND  CLOTTING
;     TITLE OF INVENTION:  FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
;     TITLE OF INVENTION:  FUSION  PROTEIN,  POLYNUCLEOTIDE AND  POLYRIBO-
;     TITLE OF INVENTION:  NUCLEOTIDE  ENCODING  THE POLYPEPTIDE, ANTI-
;     TITLE OF INVENTION:  POLYPEPTIDE ANTIBODIES, KITS  AND  METHODS OF
;     TITLE OF INVENTION:  USE THEREOF
;     NUMBER OF SEQUENCES:  5
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  V.  Amzel & Assoc.
;       STREET:  2055 No. 54550311th Broadway
;       CITY:  Walnut Creek
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94596
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS 5.0
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/607,538C
;       FILING DATE:  01-NOV-1990
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Viviana Amzel
;       REGISTRATION NUMBER:  30,930
;       REFERENCE/DOCKET NUMBER:  CRFCC-004
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (510) 943-1931
;       TELEFAX:  (510) 943-1189
;       TELEX:  N.A.
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  218 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     FRAGMENT TYPE:
US-07-607-538C-2


  Query Match           100.0%;  Score 43;  DB 1;  Length 218;
  Best Local Similarity  100.0%;  Pred. No. 0.081;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||||||||
Db     144 VQFVASYKV 152
```

```
RESULT 7
US-08-162-402B-2
; Sequence 2, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
;    APPLICANT: CERIANI, ROBERTO L.
;    APPLICANT: PETERSON, JERRY A.
;    APPLICANT: LAROCCA, DAVID J.
;    TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION: GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES: 29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Pretty, Schroeder & Poplawski
;      STREET: 444 South Flower St., 19th Floor
;      CITY: Los Angeles
;      STATE: CA
;      COUNTRY: USA
;      ZIP: 90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Diskette
;      COMPUTER: IBM Compatible
;      OPERATING SYSTEM: DOS
;      SOFTWARE: FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/162,402B
;      FILING DATE: 03-DEC-1993
;      CLASSIFICATION: 435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Amzel, Viviana
;      REGISTRATION NUMBER: 30,930
;      REFERENCE/DOCKET NUMBER: P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 213-622-7700
;      TELEFAX: 213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO: 2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 218 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: unknown
;      TOPOLOGY: unknown
;    MOLECULE TYPE: peptide
US-08-162-402B-2


  Query Match             100.0%;  Score 43;  DB 2;  Length 218;
  Best Local Similarity   100.0%;  Pred. No. 0.081;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
            |||||||||
Db     144 VQFVASYKV 152


RESULT 8
US-08-162-402B-6
; Sequence 6, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
;    APPLICANT: CERIANI, ROBERTO L.
;    APPLICANT: PETERSON, JERRY A.
;    APPLICANT: LAROCCA, DAVID J.
;    TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION: GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES: 29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Pretty, Schroeder & Poplawski
;      STREET: 444 South Flower St., 19th Floor
;      CITY: Los Angeles
;      STATE: CA
;      COUNTRY: USA
;      ZIP: 90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Diskette
;      COMPUTER: IBM Compatible
;      OPERATING SYSTEM: DOS
;      SOFTWARE: FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/162,402B
;      FILING DATE: 03-DEC-1993
;      CLASSIFICATION: 435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Amzel, Viviana
;      REGISTRATION NUMBER: 30,930
;      REFERENCE/DOCKET NUMBER: P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 213 622-7700
;      TELEFAX: 213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO: 6
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 387 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: unknown
;      TOPOLOGY: unknown
;    MOLECULE TYPE: peptide
US-08-162-402B-6


  Query Match             100.0%;  Score 43;  DB 2;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.15;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
            |||||||||
Db     313 VQFVASYKV 321


RESULT 9
US-08-162-402B-8
; Sequence 8, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
;    APPLICANT: CERIANI, ROBERTO L.
;    APPLICANT: PETERSON, JERRY A.
;    APPLICANT: LAROCCA, DAVID J.
;    TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION: GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES: 29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Pretty, Schroeder & Poplawski
;      STREET: 444 South Flower St., 19th Floor
;      CITY: Los Angeles
;      STATE: CA
;      COUNTRY: USA
;      ZIP: 90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Diskette
;      COMPUTER: IBM Compatible
;      OPERATING SYSTEM: DOS
;      SOFTWARE: FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/162,402B
;      FILING DATE: 03-DEC-1993
;      CLASSIFICATION: 435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4219
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  465 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-8


  Query Match             100.0%;  Score 43;  DB 2;  Length 465;
  Best Local Similarity   100.0%;  Pred. No. 0.18;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||||||||
Db     391 VQFVASYKV 399


RESULT  10
US-08-480-229C-21
; Sequence 21, Application US/08480229C
; Patent No. 5874562
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS-1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,229C
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8907-0026-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  321 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-480-229C-21


  Query Match             83.7%;  Score 36;  DB 2;  Length 321;
  Best Local Similarity   77.8%;  Pred. No. 3.6;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||| |||:
Db     246 VQFVGSYKL 254


RESULT  11
US-08-659-235C-21
; Sequence 21, Application US/08659235C
; Patent No. 5877281
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS 1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,235C
;      FILING DATE:  05-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8907-0034-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  321 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-659-235C-21


  Query Match             83.7%;  Score 36;  DB 2;  Length 321;
  Best Local Similarity   77.8%;  Pred. No. 3.6;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||| |||:
Db     246 VQFVGSYKL 254


RESULT  12
US-08-480-229C-10
; Sequence 10, Application US/08480229C
; Patent No. 5874562
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
```

```
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS-1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,229C
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8007-0026-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  480 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-229C-10


  Query Match             83.7%;  Score 36;  DB 2;  Length 480;
  Best Local Similarity   77.8%;  Pred. No. 5.5;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
             |||| |||:
Db     402 VQFVGSYKL 410


RESULT  14
US-08-659-235C-10
; Sequence 10, Application US/08659235C
; Patent No. 5877281
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS-1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,235C
;      FILING DATE:  05-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8007-0034-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  480 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-659-235C-10


  Query Match             83.7%;  Score 36;  DB 2;  Length 480;
  Best Local Similarity   77.8%;  Pred. No. 5.5;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
             |||| |||:
Db     402 VQFVGSYKL 410


RESULT  14
US-08-480-229C-14
; Sequence 14, Application US/08480229C
; Patent No. 5874562
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS-1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,229C
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8007-0026-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  513 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-229C-14


  Query Match             83.7%;  Score 36;  DB 2;  Length 513;
```

```
  Best Local Similarity   77.8%;  Pred. No. 5.9;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||| ||||:
Db     435 VQFVGSYKL 443


RESULT  15
US-08-659-235C-14
; Sequence 14, Application US/08659235C
; Patent No. 5877281
;  GENERAL INFORMATION:
;    APPLICANT:  Quertermous, Thomas
;    APPLICANT:  Hogan, Brigid
;    APPLICANT:  Snodgrass, H. Ralph
;    APPLICANT:  Zupancic, Thomas J.
;    TITLE OF INVENTION:  DEVELOPMENTALLY-REGULATED ENDOTHELIAL
;    TITLE OF INVENTION:  CELL LOCUS-1
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,235C
;      FILING DATE:  05-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poissant, Brian M.
;      REGISTRATION NUMBER:  28,462
;      REFERENCE/DOCKET NUMBER:  8907-0034-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 Pennie
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  513 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-659-235C-14


  Query Match             83.7%;  Score 36;  DB 2;  Length 513;
  Best Local Similarity   77.8%;  Pred. No. 5.9;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||| ||||:
Db     435 VQFVGSYKL 443


Search completed: September  5, 2002, 15:27:36
Job time: 321 sec
```

OM protein - protein search, using sw model

```
Run on:          September 5, 2002, 15:26:31 ; Search time 158.43 Seconds
                                               (without alignments)
                                               6.310 Million cell updates/sec

Title:           US-09-744-804-40
Perfect score:   43
Sequence:        1 VQFVASYKV 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters:      747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_032802:*
                 1:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
                 2:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
                 3:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
                 4:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
                 5:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
                 6:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
                 7:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
                 8:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
                 9:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
                 10:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA19[illegible].DAT:*
                 11:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
                 12:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
                 13:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
                 14:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
                 15:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
                 16:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
                 17:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
                 18:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
                 19:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
                 20:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
                 21:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
                 22:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	100.0	9	21	AAY82845	Lactadherin (BA-46
2	43	100.0	85	20	AAW94690	Discoidin I/factor
3	43	100.0	217	16	AAR77253	HMFG 46 kDa antige
4	43	100.0	218	16	AAR77254	HMFG 46 kDa antige
5	43	100.0	387	16	AAR77252	HMFG 46 kDa antige
6	43	100.0	387	21	AAY94453	Human lactadherin
7	36	83.7	52	22	ABB44260	Peptide #11766 enc
8	36	83.7	52	22	ABB27131	Protein #9130 enco
9	36	83.7	52	22	AAM58323	Human brain expres
10	36	83.7	52	22	AAM65298	Human brain expres
11	36	83.7	52	22	AAM77997	Human bone marrow
12	36	83.7	52	22	AAM21889	Peptide #8323 enco
13	36	83.7	52	22	AAM38215	Peptide #12252 enc
14	36	83.7	321	20	AAW91697	Human milk fat glo
15	36	83.7	480	18	AAW10364	Mouse developmenta
16	36	83.7	480	20	AAW94683	Murine Del-1 prote
17	36	83.7	481	18	AAW10365	Human developmenta
18	36	83.7	513	20	AAW94685	Human Del-1 protei
19	34	79.1	320	20	AAW98891	Peptide SEQ ID NO:
20	34	79.1	426	21	AAY94454	Mouse lactadherin
21	33	76.7	76	16	AAR87676	(24-Arg) MCP-1. H
22	33	76.7	533	20	AAY04367	Methanococcus mari
23	32	74.4	483	22	ABB65653	Drosophila melanog
24	32	74.4	610	22	ABB68385	Drosophila melanog
25	32	74.4	663	21	AAB18996	A partial core 2 b
26	31	72.1	279	21	AAB16698	Bacteriophage Dp-1
27	30	69.8	134	22	ABB62376	Drosophila melanog
28	30	69.8	280	21	AAG15435	Arabidopsis thalia
29	30	69.8	281	21	AAG36456	Arabidopsis thalia
30	30	69.8	281	21	AAG43413	Arabidopsis thalia
31	30	69.8	286	21	AAG36455	Arabidopsis thalia
32	30	69.8	286	21	AAG43412	Arabidopsis thalia
33	30	69.8	287	21	AAG15434	Arabidopsis thalia
34	30	69.8	290	21	AAG15433	Arabidopsis thalia
35	30	69.8	296	21	AAG36454	Arabidopsis thalia
36	30	69.8	296	21	AAG43411	Arabidopsis thalia
37	30	69.8	573	22	AAB70846	Tobacco PRPP-amido
38	30	69.8	2162	22	ABB60750	Drosophila melanog
39	29	67.4	85	18	AAW10367	Del-1 discoidin I-
40	29	67.4	85	20	AAW94689	Discoidin I/factor
41	29	67.4	85	20	AAW94693	Discoidin I/factor
42	[illegible]	[illegible]	[illegible]	[illegible]	AAW[illegible]	[illegible] I/factor
43	[illegible]	[illegible]	[illegible]	[illegible]	AAW[illegible]	[illegible] I/factor
44	[illegible]	[illegible]	184	[illegible]	AAY82[illegible]	Human [illegible] 1 doma
45	29	67.4	184	21	AAY82850	Rhesus [illegible] 1 dom

ALIGNMENTS

```
RESULT 1
AAY82845
ID   AAY82845 standard; peptide; 9 AA.
XX
AC   AAY82845;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Lactadherin (BA-46) peptide fragment (tumour associated antigen).
XX
KW   Tumour associated antigen peptide; TAA; cancer; carcinoma;
KW   treatment; prevention; cure; anti-tumour vaccine; metastases;
KW   [illegible]
KW   stomach; carcinoma; MHC Class I; HLA-A2; human;
KW   Major Histocompatibility Complex; [illegible];
KW   prostate specific antigen; prostate specific membrane antigen;
KW   prostate acid phosphatase; mucin; lactadherin;
KW   teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRIPTO-1
XX
OS   Homo sapiens.
XX
PN   WO200006723-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-IL00417.
XX
PR   30-JUL-1998;   98IL-0125608.
XX
PA   (YEDA ) YEDA RES & DEV CO LTD.
PA   (BIOT ) BIO TECHNOLOGY GEN CORP.
XX
PI   Eisenbach L,  Carmon L,  Tirosh B,  Bar haim E,  Paz A,  Fridkin M;
PI   Fitzer-attas C;
```

XX
DR WPI; 2000-205463/18.
XX
PT Tumor associated antigen peptides, especially derived from uroplakin,
PT useful as vaccines to prevent or cure cancers including breast,
PT bladder, prostate, pancreas, ovary, thyroid, colon and stomach
XX
PS Claim 17; Page 100; 113pp; English.
XX
CC Tumour associated antigen peptides (TAA) may be used for the
CC treatment, prevention and cure of cancer or cancer metastases. The
CC cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
CC colon, stomach, head or neck cancer or a carcinoma. The tumour
CC associated antigens are presentable to the immune system by HLA A2
CC molecules and are generally between 8 to 10 amino acids in length.
CC The amino acids located at positions 2 and 9 of the tumour associated
CC antigens are the anchor residues which participate in the binding to
CC MHC class I molecules, more specifically HLA-A2. More tumour
CC associated antigens are described in GENESEQ records AAY82806-Y82882.
CC Those tumour associated antigens described in records AAY82806-Y82824
CC and AAY82855-YY82869 are derived from Uroplakin, such as Uroplakin II,
CC Uroplakin Ia, Uroplakin III and Uroplakin Ib. Those described in
CC records AAY82825-Y82829 are derived from prostate specific antigen
CC (PSA). Those described in records AAY82830-Y82835 are derived from
CC prostate specific membrane antigen (PSMA). Those described in
CC records YY82836 AAY82839 are derived from prostate acid phosphotase
CC (PAP). Those described in records AAY82840-Y82846 are derived
CC from Lactadherin (BA-46). Those described in records AAY82847-Y82854
CC are derived from Mucin and those described in records AAY82871-Y82882
CC are derived from Teratocarcinoma derived growth factor (CRIPTO-1).
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 43; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VQFVASYKV 9
||||||||| |
Db 1 vqtvasykv 9

RESULT 2
AAW94690
ID AAW94690 standard; peptide; 85 AA.
XX
AC AAW94690;
XX
DT 05-MAY-1999 (first entry)
XX
DE Discoidin I/factor VIII like domain #2.
XX
KW Del-1; developmentally regulated endothelial cell locus 1; cancer;
KW discoidin I, factor VIII like domain, epidermal growth factor; EGF;
KW diabetic retinopathy; rheumatoid arthritis; endometriosis;
KW angiogenesis.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US5877281-A.
XX
PD 02-MAR-1999.
XX
PF 05-JUN-1996; 96US-0659235.
XX
PR 05-JUN-1996; 96US-0659235.
PR 07-JUN-1995; 95US-0480229.
XX
PA (PROG-) PROGENITOR INC.
PA (UYVA-) UNIV VANDERBILT.
XX
PI Hogan B, Quertermous T, Snodgrass HR, Zupancic TJ;
XX
DR WPI; 1999-189720/16.
XX
PT Del 1 polypeptide sequences useful for treatment of cancer,
PT diabetic retinopathy, rheumatoid arthritis and endometriosis
XX
PS Example; Column 39-40; 73pp; English.
XX
CC The present sequence represents a discoidin I/factor VIII like domain.
CC The present invention describes developmentally-regulated endothelial
CC cell locus 1 (Del 1). Del 1 has epidermal growth factor like domains and
CC discoidin I/factor VIII-like domains. The Del 1 proteins have an
CC inhibitory effect on angiogenesis (blood vessel growth), this activity
CC may be useful clinically to prevent neovascularisation of tissues such
CC as tumour nodules and prevention of metastases. The anti-angiogenic
CC activity of Del 1 may be used to treat abnormal conditions that result
CC from angiogenesis, including cancer, diabetic retinopathy, rheumatoid
CC arthritis and endometriosis. Since Del-1 promotes angiogenesis it can be
CC used to treat cardiac ischaemia, thrombotic stroke, wound healing and
CC peripheral vascular disease. Del 1 is also useful for promoting bone
CC formation. Del-1 binds to alpha V beta 3 integrin, and is an apoptosis
CC inducer.
XX
SQ Sequence 85 AA;

Query Match 100.0%; Score 43; DB 20; Length 85;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VQFVASYKV 9
||||||||| |
Db 24 vqfvasykv 32

RESULT 3
AAR77253
ID AAR77253 standard; Protein; 217 AA.
XX
AC AAR77253;
XX
DT 21-NOV-1995 (first entry)
XX
DE HMFG 46 kDa antigen C terminal region.
XX
KW HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
KW epithelium; tumor; breast cancer; monoclonal antibody; MAb.
XX
OS Homo sapiens.
XX
PN WO9515171-A.
XX
PD 08-JUN-1995.
XX
PF 05-DEC-1994; 94WO-US13967.
XX
PR 03-DEC-1993; 93US-0162402.
XX
PA (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI Ceriani RL, Larocca DJ, Peterson JA;
XX
DR WPI; 1995-215151/28.
XX
PT 46 kD apparent molecular weight human milk fat globule antigen -
PT used in assays to determine the presence of a cancerous tumour of
PT epithelial origin, and in a vaccine against neoplastic tumours
XX
PS Claim 7; Page 41; 68pp; English.
XX
CC A partial cDNA clone BA46-1 (AAQ91199) for the 46 kDa HMFG antigen,
CC a major component of the apical surface of the normal breast

```
CC   epithelial cell, was obtained by PCR of cDNA from a lactating
CC   breast cDNA library.  The C-terminal region of the encoded protein
CC   showed 43% identity to corresp  regions of human Factor-V and 38%
CC   to Factor-VIII.
XX
SQ   Sequence   217 AA;


  Query Match             100.0%;  Score 43;  DB 16;  Length 217;
  Best Local Similarity   100.0%;  Pred  No  0.33;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 VQFVASYKV 9
          |||||||||
Db    143 vqfvasykv 151


RESULT   4
AAR77254
ID   AAR77254 standard, Protein, 218 AA.
XX
AC   AAR77254;
XX
DT   21-NOV-1995  (first entry)
XX
DE   HMFG 46 kDa antigen partial sequence.
XX
KW   HMFG; human milk fat globule, antigen; immunogen, vaccine, virucide,
KW   epithelium; tumor, breast cancer; monoclonal antibody; MAb.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   69..71
FT                   /label= N-glycosylation_site
FT   Modified-site   156..158
FT                   /label= N-glycosylation_site
FT   Modified-site   160..162
FT                   /label= N-glycosylation_site
FT   Modified-site   181..183
FT                   /label= N-glycosylation_site
XX
PN   WO9515171-A.
XX
PD   08-JUN-1995
XX
PF   05-DEC-1994;   94WO-US13967
XX
PP   03-DEC-1993;   93US-0162402.
XX
PA   (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI   Ceriani RL,  Larocca DJ,  Peterson JA;
XX
DR   WPI; 1995-215151/28
DR   N-PSDB; AAQ91199.
XX
PT   46 kD apparent molecular weight human milk fat globule antigen -
PT   used in assays to determine the presence of a cancerous tumour of
PT   epithelial origin, and in a vaccine against neoplastic tumours
XX
PS   Example 7; Page 38-40; 68pp; English.
XX
CC   A partial cDNA clone BA46-1 (AAQ91199) for the 46 kDa HMFG antigen,
CC   a major component of the apical surface of the normal breast
CC   epithelial cell, was obtained by PCR of cDNA from a lactating
CC   breast cDNA library.
XX
SQ   Sequence   218 AA;


  Query Match             100.0%;  Score 43;  DB 16;  Length 218;
  Best Local Similarity   100.0%;  Pred  No  0.34;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 VQFVASYKV 9
          |||||||||
Db    144 vqfvasykv 152


RESULT   5
AAR77252
ID   AAR77252 standard; Protein; 387 AA.
XX
AC   AAR77252;
XX
DT   21-NOV-1995  (first entry)
XX
DE   HMFG 46 kDa antigen.
XX
KW   HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
KW   epithelium, tumor, breast cancer, monoclonal antibody; MAb.
XX
OS   Homo sapiens.
XX
PN   WO9515171-A.
XX
PD   08-JUN-1995.
XX
PF   05-DEC-1994;   94WO-US13967.
XX
PP   03-DEC-1993;   93US-0162402.
XX
PA   (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI   Ceriani RL,  Larocca DJ,  Peterson JA;
XX
DR   WPI; 1995-215151/28
DR   N-PSDB; AAQ91198.
XX
PT   46 kD apparent molecular weight human milk fat globule antigen -
PT   used in assays to determine the presence of a cancerous tumour of
PT   epithelial origin, and in a vaccine against neoplastic tumours
XX
PS   Claim 6; Page 46-47; 68pp; English.
XX
CC   A complete cDNA sequence for the 46 kDa HMFG antigen, a major
CC   component of the apical surface of the normal breast epithelial
CC   cell, was obtained by PCR and RACE methods.  cDNA clones can be
CC   used to prepare MAbs for use in immunotherapy, immunohistopathology,
CC   prognosis, diagnosis, imaging and therapy.  Recombinant antigen can
CC   be expressed in prokaryotic or (glycosylated) in eukaryotic cells.
XX
SQ   Sequence   387 AA;


  Query Match             100.0%;  Score 43;  DB 16;  Length 387;
  Best Local Similarity   100.0%;  Pred  No  0.61;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 VQFVASYKV 9
          |||||||||
Db    313 vqfvasykv 321


RESULT   6
AAY94453
ID   AAY94453 standard; Protein; 387 AA.
XX
AC   AAY94453;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Human lactadherin protein.
XX
KW   Human; lactadherin; MGF-E8; anti-tumour; immune response;
```

```
KW   exosome; dendritic cell.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Secretion_signal
FT   Protein         24..387
FT                   /label= Lactadherin
FT   Binding-site    46..48
FT                   /label= Integrin_binding_site
XX
PN   EP1004664-A1.
XX
PD   31-MAY-2000.
XX
PF   24-NOV-1998;    98EP-0402925.
XX
PR   24-NOV-1998;    98EP-0402925.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (CURI-) INST CURIE.
XX
DR   WPI; 2000-352597/31.
DR   N-PSDB; AAA27140.
XX
PT   Chimeric isolated (human) lactadherin polypeptide that functions as an
PT   adaptor of cross-priming to eliminate pathogenic antigens, e.g. in
PT   tumors -
XX
PS   Example 3; Page 12; 20pp; English.
XX
CC   Lactadherin protein was found in exosomes produced by dendritic cells.
CC   The protein is involved in the phagocytosis of particulate antigens by
CC   dendritic cells. Exosomes produced by dendritic cells exposed to
CC   tumour antigens induce potent immune responses. Lactadherin or variants
CC   of it may be used in the mediation of an immune response. Variants of
CC   lactadherin may be used for inhibition and/or stimulation of the
CC   cross priming of antigens and stimulation of the phagocytosis of
CC   antigens by dendritic cells. Compositions derived from lactadherin can
CC   also be used to monitor an immune response, more specifically a CTL
CC   (cytotoxic T-lymphocyte) response and also to produce CTLs specific
CC   for a selected antigen. The present sequence is the human
CC   lactadherin protein.
XX
SQ   Sequence   387 AA;


  Query Match             100.0%;  Score 43;  DB 21;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.61;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||||||||
Db     313 vqfvasykv 321


RESULT   7
ABB44260
ID   ABB44260 standard; Peptide; 52 AA.
XX
AC   ABB44260;
XX
DT   04-FEB-2002  (first entry)
XX
DE   Peptide #11766 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression, single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277 A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024264.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 27; SEQ ID NO 36895; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a peptide encoded by a single exon
CC   nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   52 AA;


  Query Match             83.7%;  Score 36;  DB 22;  Length 52;
  Best Local Similarity   77.8%;  Pred. No. 2.2;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||| |||:
Db      23 vqfvqsykl 31


RESULT   8
ABB27131
ID   ABB27131 standard; Protein; 52 AA.
XX
AC   ABB27131;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Protein #9130 encoded by probe for measuring heart cell gene expression.
XX
KW   Human; gene expression, heart, microarray, vascular system;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
```

PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 15, SEQ ID NO 28991, 550pp, English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535 ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease
CC Note. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 52 AA;

Query Match 83.7%; Score 36; DB 22; Length 52;
Best Local Similarity 77.8%; Pred. No. 2.2;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VQFVASYKV 9
|||| |||:
Db 23 vqfvgsykl 31

RESULT 9
AAM58323
ID AAM58323 standard; Protein; 52 AA.
XX
AC AAM58323;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30428.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00567.
XX
PR 04-FEB-2000; 2000US-0180412.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4, SEQ ID NO. 30428, 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 52 AA;

Query Match 83.7%; Score 36; DB 22; Length 52;
Best Local Similarity 77.8%; Pred. No. 2.2;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VQFVASYKV 9
|||| |||:
Db 23 vqfvgsykl 31

RESULT 10
AAM65298
ID AAM65298 standard; Protein; 52 AA.
XX
AC AAM65298;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 37403.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00567.
XX
PR 04-FEB-2000; 2000US-0180412.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4, SEQ ID NO. 37403, 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of

```
CC   the probes of the invention.
XX
SQ   Sequence   52 AA;


  Query Match             83.7%;  Score 36;  DB 22;  Length 52;
  Best Local Similarity   77.8%;  Pred. No. 2.2;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||| ||| :
Db      23 vqfvqsykl 31


RESULT  11
AAM77997
ID   AAM77997 standard; Protein; 52 AA.
XX
AC   AAM77997;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed probe encoded protein SEQ ID NO: 38303.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS   Homo sapiens
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668
XX
PR   04-FEB-2000; 2000US-0180312
PR   26-MAY-2000; 2000US-0207456
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4, SEQ ID NO. 38303, 658pp - Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is a
CC   protein encoded by one of the probes of the invention
XX
SQ   Sequence   52 AA;


  Query Match             83.7%;  Score 36;  DB 22;  Length 52;
  Best Local Similarity   77.8%;  Pred. No. 2.2;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||| ||| :
Db      23 vqfvqsykl 31


RESULT  12
AAM21889
ID   AAM21889 standard; Protein; 52 AA.
XX
AC   AAM21889;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Peptide #8323 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 27, SEQ ID No 26715; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP, see AA110368 AA128459). The present sequence is a peptide encoded
CC   by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC   can be used to produce a single exon microarray, which can be used for
CC   measuring human gene expression in a sample derived from human cervical
CC   epithelial cells. By measuring gene expression, the probes are therefore
CC   useful in grading and/or staging of diseases of the cervix, notably
CC   cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   52 AA;


  Query Match             83.7%;  Score 36;  DB 22;  Length 52;
  Best Local Similarity   77.8%;  Pred. No. 2.2;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||| ||| :
Db      23 vqfvqsykl 31


RESULT  13
AAM38215
ID   AAM38215 standard; Protein; 52 AA.
XX
AC   AAM38215;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Peptide #12252 encoded by probe for measuring placental gene expression.
```

```
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632766.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024063.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53
XX
PT   Human genome derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta
XX
PS   Claim 27; SEQ ID No 38484; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP:
CC   see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders.
XX
SQ   Sequence   52 AA;


  Query Match            83.7%;  Score 36;  DB 22;  Length 52;
  Best Local Similarity   77.8%;  Pred. No. 2.2;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
            |||| |||:
Db      23 vqfvqsykl 31


RESULT  14
AAW94697
ID   AAW94697 standard; Protein; 721 AA.
XX
AC   AAW94697;
XX
DT   05-MAY-1999  (first entry)
XX
DE   Human milk fat globule protein MFG-E8.
XX
KW   Del-1; developmentally regulated endothelial cell locus 1, cancer,
KW   discoidin I; factor VIII-like domain; epidermal growth factor; EGF;
KW   diabetic retinopathy; rheumatoid arthritis, endometriosis,
KW   angiogenesis.
XX
OS   Homo sapiens.
XX
PN   US5877281-A.
XX
PD   02-MAR-1999.
XX
PF   05-JUN-1996;   96US-0659235.
XX
PR   05-JUN-1996;   96US-0659235.
PR   07-JUN-1995;   95US-0480229.
XX
PA   (PROG-) PROGENITOR INC.
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Hogan B,  Quertermous T,  Snodgrass HR,  Zupancic TJ;
XX
DR   WPI; 1999-189720/16.
XX
PT   Del 1 polypeptide sequences    useful for treatment of cancer,
PT   diabetic retinopathy, rheumatoid arthritis and endometriosis
XX
PS   Example; Column 63-64; 74pp; English.
XX
CC   The present sequence represents human milk fat globule protein MFG-E8,
CC   which has homology to the developmentally-regulated endothelial cell
CC   locus 1 (Del 1). The Del 1 protein has epidermal growth factor (EGF) like
CC   domains and discoidin I/factor VIII like domains. The Del-1 proteins
CC   have an inhibitory effect on angiogenesis (blood vessel growth), this
CC   activity may be useful clinically to prevent neovascularisation of
CC   tissues such as tumour nodules and prevention of metastases. The anti-
CC   angiogenic activity of Del 1 may be used to treat abnormal conditions
CC   that result from angiogenesis, including cancer, diabetic retinopathy,
CC   rheumatoid arthritis and endometriosis. Since Del-1 promotes
CC   angiogenesis it can be used to treat cardiac ischaemia, thrombotic
CC   stroke, wound healing and peripheral vascular disease. Del-1 is also
CC   useful for promoting bone formation. Del 1 binds to alpha V beta 3
CC   integrin, and is an apoptosis inducer.
XX
SQ   Sequence   321 AA;


  Query Match            83.7%;  Score 36;  DB 20;  Length 321;
  Best Local Similarity   77.8%;  Pred. No. 14;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
            |||| |||:
Db     246 vqfvqsykl 254


RESULT  15
AAW10364
ID   AAW10364 standard; Protein; 480 AA.
XX
AC   AAW10364;
XX
DT   03-MAY-1997  (first entry)
XX
DE   Mouse developmentally-regulated endothelial cell locus-1 protein.
XX
KW   Del 1; developmentally regulated endothelial cell locus-1;
KW   signal transduction; cancer; tumour marker; angiogenesis;
KW   diabetic retinopathy, rheumatoid arthritis, endometriosis;
KW   cardiac ischaemia; stroke; vascular disease; wound healing;
KW   vulnerary; bone formation; diagnosis; therapy.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1..23
FT                    /label= Sig_peptide
FT   Protein          24..480
FT                    /label= Mat_protein
FT   Domain           26..59
FT                    /label= EGF-1
FT                    /note= "epidermal growth factor-like domain 1"
FT   Domain           78..116
FT                    /label= EGF-2
FT                    /note= "epidermal growth factor-like domain 2"
FT   Domain           123..154
FT                    /label= EGF-3
```

```
FT                      /note "epidermal growth factor-like domain 3"
FT   Domain             158..314
FT                      /label= Discoidin-1
FT                      /note= "discoidin I/factor VIII-like domain 1"
FT   Domain             319..476
FT                      /label= Discoidin-2
FT                      /note= "discoidin I/factor VIII-like domain 2"
XX
PN   WO9640769-A1.
XX
PD   19-DEC-1996.
XX
PF   05-JUN-1996;   96WO-US09456.
XX
PR   07-JUN-1995;   95US-0480229.
XX
PA   (PROG-) PROGENITOR INC.
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Hogan B, Quertermous T, Snodgrass HR, Zupancic TJ;
XX
DR   WPI; 1997-052233/05.
DR   N-PSDB; AAT47338.
XX
PT   New developmentally regulated endothelial cell locus-1 (del 1) gene
PT   - used to develop prods. for the diagnosis and treatment of cancer
PT   and conditions involving abnormal angiogenesis
XX
PS   Claim 3; Fig 6; 137pp; English.
XX
CC   Murine Del-1 (AAW10364) is the polypeptide product of the murine
CC   developmentally-regulated endothelial cell locus-1 (del-1) gene
CC   (AAT47338).  It shows 94% amino acid homology to the human Del-1
CC   protein (AAW10365).  Structurally, members of this novel gene family
CC   contain 3 EGF-like domains and 2 discoidin I/factor VIII-like
CC   domains.  Del 1 is expressed in endothelial and certain tumour
CC   cells.  Its ability to inhibit vascular formation allows its used
CC   as an anti-angiogenic agent.  It can be used as a tumour marker,
CC   to identify Del-1 binding partners, and to modulate endothelial
CC   cell growth and blood vessel formation.  Recombinant Del-1 can be
CC   produced in transformed host cells utilising vectors incorporating
CC   del-1 nucleic acids.
XX
SQ   Sequence   480 AA;


  Query Match             83.7%;  Score 36;  DB 18;  Length 480;
  Best Local Similarity   77.8%;  Pred. No. 22;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||| |||:
Db     402 vqfvqsykl 410



Search completed: September  5, 2002, 15:26:32
Job time: 372 sec
```

GenCore version 4.5

OM protein protein search, using sw model

Run on: September 5, 2002, 15:28:53 , Search time 69.78 Seconds
(without alignments)
12.393 Million cell updates/sec

Title: US-09-744-804-40
Perfect score: 43
Sequence: 1 VQFVASYKV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters. 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	43	100.0	218	2	A47285	milk fat globule p
2	39	90.7	427	2	JC4915	ags protein precur
3	36	83.7	409	2	T11743	pP47 protein - pig
4	34	79.1	463	1	A36479	milk fat globule m
5	33	76.7	401	2	S65138	glycoprotein antig
6	33	76.7	427	2	S74311	PAS 6/7 protein pr
7	33	76.7	533	2	T46975	lysine--tRNA ligas
8	32	74.4	445	2	T43420	probable protein k
9	32	74.4	958	2	S15566	pol protein simi
10	32	74.4	1161	2	S18738	pol protein simi
11	31	72.1	213	2	B82260	rare lipoprotein B
12	31	72.1	217	2	T40730	probable inorganic
13	31	72.1	308	2	B82789	hypothetical prote
14	31	72.1	344	2	T08110	integral membrane
15	31	72.1	489	1	SYRSYS	glutamate--tRNA li
16	31	72.1	1226	2	HE1479	protoporphyrin IX
17	30	69.8	260	2	G83686	hypothetical prote
18	30	69.8	290	2	G85040	probable xyloglucan
19	30	69.8	296	2	G49530	xyloglucan endo-1,
20	30	69.8	345	2	AB1392	yeast translation
21	30	69.8	374	2	A82588	membrane bound lyt
22	30	69.8	396	2	B75175	myo-inositol 1 pho
23	30	69.8	476	2	JC7189	tubulointerstitial
24	30	69.8	478	2	AF0153	6 phospho beta glu
25	30	69.8	529	2	T20121	hypothetical prote
26	30	69.8	576	2	T25740	hypothetical prote
27	30	69.8	641	2	B82884	DNA primase UL194
28	30	69.8	672	2	T25874	hypothetical prote
29	30	69.8	1216	2	T09224	spindle assembly c
30	29	67.4	106	2	TI7951	protein disulfide-
31	29	67.4	149	1	DLDOID	discoidin I chain
32	29	67.4	149	2	B03482	discoidin I chain
33	29	67.4	187	2	AF3009	hypothetical prote
34	29	67.4	195	2	B98274	hypothetical prote
35	29	67.4	253	1	DLDOIA	discoidin I chain
36	29	67.4	253	1	DLDOIC	discoidin I chain
37	29	67.4	277	1	QRMSST	sulfate/thiosulfat
38	29	67.4	277	2	G91040	sulfate transport
39	29	67.4	277	2	AB0966	sulfate transport
40	29	67.4	277	2	B85885	sulfate transport
41	29	67.4	277	2	AH0811	sulphate transport
42	29	67.4	292	2	A49539	xyloglucan endo-1,
43	29	67.4	297	2	T10523	xyloglucan endo-1,
44	29	67.4	338	2	T26873	hypothetical prote
45	29	67.4	345	2	AC1767	yeast translation

ALIGNMENTS

RESULT 1
A47285
milk fat globule protein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-May-1999
C;Accession: A47285
R;Larocca, D., Peterson, J.A., Urrea, R., Kuniyoshi, J., Bistrain, A.M.; Ceriani, R.L.
Cancer Res. 51, 4994-4998, 1991
A;Title: A M : 46,000 human milk fat globule protein that is highly expressed in huma
A;Reference number: A47285; MUID:91371351
A;Accession: A47285
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-218 <LAR>
A;Cross-references: GB:S56151; NID:g235396; PIDN:AAB19771.1; PID:g235397
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
F;1-56/Domain: discoidin I amino-terminal homology (fragment) <DN1>
F;60-218/Domain: discoidin I amino-terminal homology <DN2>

Query Match 100.0%; Score 43; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 0.067;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VQFVASYKV 9
||||||||| |
Db 144 VQFVASYKV 152

RESULT 2
JC4915
ags protein precursor - rat
N;Alternate names: O-acetyl GD3 ganglioside
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
C;Accession: JC4915
R;Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A;Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
A;Reference number: JC4915; MUID:96274422
A;Accession: JC4915
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-427 <OGU>
A;Cross-references: EMBL:D84048; NID:g1620806; PIDN:BAA12216.1; PID:g1620807
A;Experimental source: CST cell
C;Comment: This protein is required for the O-acetylation of GD3 ganglioside finali
C;Genetics:
A;Gene: ags
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
F;1-21/Domain: signal sequence #status predicted <SIG>
F;28-66/Domain: EGF homology <EGF>

```
F;68-107/Domain: EGF homology <EG2>
F;110-267/Domain: discoidin I amino terminal homology <DN1>
F;271-427/Domain: discoidin I amino terminal homology <DN2>


  Query Match             90.7%;  Score 39;  DB 2;  Length 427;
  Best Local Similarity   77.8%;  Pred. No. 1;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
            :|:||||||
Db     455 IQYVASYKV 363


RESULT    3
T11743
pP47 protein - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21 Jul 2000
C;Accession: T11743
R;Ensslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toepfe
Biol. Reprod. 58, 1057-1064, 1998
A;Title: Molecular cloning and characterization of P47, a novel boar sperm associated ze
A;Reference number: A17325; MUID:98206817
A;Accession: T11743
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-409 <ENS>
A;Cross-references: EMBL:Y11683; NID:g2652927; PIDN:CAA72379.1; PID:g2652928
A;Experimental source: testis
C;Function:
A;Description: may be involved in membrane remodeling and/or function as a zona pellucid
C;Superfamily: unassigned EGF-related proteins; EGF homology
F;6-40/Domain: EGF homology <EGF>


  Query Match             83.7%;  Score 36;  DB 2;  Length 409;
  Best Local Similarity   66.7%;  Pred. No. 4.7;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
            :|:|:|||
Db     345 IQYVAAYKV 343


RESULT    4
A36479
milk fat globule membrane protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A36479
R;Stubbs, J.D.; Lekutis, C.; Singer, K.L., Bui, A.; Yuzuki, D.; Srinivasan, U.; Parry, G
Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
A;Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the exi
A;Reference number: A36479; MUID:91046008
A;Accession: A36479
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-463 <STU>
A;Cross-references: GB:M38337; NID:g199142; PIDN:AAA39534.1; PID:g199143
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
C;Keywords: membrane protein
F;28-60/Domain: EGF homology <EG1>
F;68-107/Domain: EGF homology <EG2>
F;147-303/Domain: discoidin I amino terminal homology <DN1>
F;307-463/Domain: discoidin I amino terminal homology <DN2>


  Query Match             79.1%;  Score 34;  DB 1;  Length 463;
  Best Local Similarity   66.7%;  Pred. No. 15;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
            :|:| ||||
Db     331 IQYVESYKV 399


RESULT    5
S65138
glycoprotein antigen MGP57/53, mammary gland - bovine (fragment)
N;Alternate names: glycoprotein <MAJ>; 52/major fat globule membrane protein; MFG E
C;Species: Bos primigenius taurus (cattle)
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
C;Accession: S65138; G48394
R;Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A;Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal
A;Reference number: S65138; MUID:96125736
A;Accession: S65138
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-401 <AOK>
R;Mather, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A;Title: The major fat-globule membrane proteins, bovine components 15/16 and quinea
II-like sequences.
A;Reference number: A48394; MUID:93250576
A;Accession: G48394
A;Status: preliminary
A;Molecule type: protein
A;Residues: 207-220 <MAT>
A;Experimental source: milk
A;Note: sequence extracted from NCBI backbone (NCBIP:131457)
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
C;Keywords: glycoprotein
F;1-32/Domain: EGF homology (fragment) <EG1>
F;40-79/Domain: EGF homology <EG2>
F;82-239/Domain: discoidin I amino-terminal homology <DN1>
F;243-401/Domain: discoidin I amino terminal homology <DN2>


  Query Match             76.7%;  Score 33;  DB 2;  Length 401;
  Best Local Similarity   55.6%;  Pred. No. 21;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
            :|:||:|:|
Db     327 IQYVAAYRV 335


RESULT    6
S74211
PAS-6/7 protein precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000
C;Accession: S74211; S78114; S24181
R;Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
Eur. J. Biochem. 240, 628-636, 1996
A;Title: Characterization of glycoprotein PAS 6/7 from membranes of bovine milk fat g
A;Reference number: S74211; MUID:97008954
A;Accession: S74211
A;Molecule type: mRNA
A;Residues: 1-427 <HVA>
A;Cross-references: EMBL:X91895; NID:g1642779; PIDN:CAA62997.1; PID:g1642779
A;Accession: S78114
A;Molecule type: protein
A;Residues: 29-85,96-110,148-165,174-216,221-242,248-277,285-293,309-337,339-420,425
R;Kim, D.H.; Kanno, C.; Mizokami, Y.
Biochim. Biophys. Acta 1122, 203-211, 1992
A;Title: Purification and characterization of major glycoproteins, PAS 6 and PAS 7, f
A;Reference number: S23926; MUID:92353107
A;Accession: S24181
A;Molecule type: protein
A;Residues: 383-394 <KIM>
C;Superfamily: milk fat globule protein; discoidin I amino terminal homology; EGF hom
C;Keywords: blocked amino end; disulfide bond; glycoprotein; milk
```

```
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-427/Product: FAS 6/7 protein #status experimental <MAT>
F;24-58/Domain: EGF homology <EG1>
F;66-105/Domain: EGF homology <EG2>
F;108-265/Domain: discoidin I amino-terminal homology <DN1>
F;269-427/Domain: discoidin I amino-terminal homology <DN2>
F;24-35,29-47,49-58,66-77,71-94,96-105/Disulfide bonds. #status predicted
F;27/Binding site: carbohydrate (Ser) (covalent) #status experimental
F;34/Binding site: carbohydrate (Thr) (covalent) #status experimental
F;59,227/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;109-265,252-256,270-427/Disulfide bonds. #status experimental

  Query Match             76.7%;  Score 33;  DB 2;  Length 427;
  Best Local Similarity   55.6%;  Pred. No. 23;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VQFVASYKV 9
              :|:||:|:|
Db        353 IQYVAAYPV 361

RESULT 7
T46975
lysine--tRNA ligase (EC 6.1.1.6) [validated] - Methanococcus maripaludis
N;Alternate names: lysyl-tRNA synthetase
C;Species: Methanococcus maripaludis
C;Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C;Accession: T46975
R;Ibba, M.; Morgan, S.; Curnow, A.W.; Pridmore, D.R.; Vothknecht, U.C.; Gardner, W.; Lin
Science 278, 1119-1122, 1997
A;Title: A euryarchaeal lysyl-tRNA synthetase: resemblance to class I synthetases.
A;Reference number: A58679; MUID:98016282
A;Accession: T46975
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-533 <IBB>
A;Cross-references: EMBL:AF009824; NID:g2645488; PIDN:AAB87410.1; PID:g2645489
A;Experimental source: strain JJ
C;Genetics:
A;Gene: lysS
C;Function:
A;Description: EC 6.1.1.6 [validated, MUID:98016282]
C;Superfamily: Lyme disease spirochete lysine--tRNA ligase
C;Keywords: ligase

  Query Match             76.7%;  Score 33;  DB 2;  Length 533;
  Best Local Similarity   85.7%;  Pred. No. 29;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 FVASYKV 9
              ||||||:
Db        497 FVASYKI 503

RESULT 8
T43420
probable protein kinase (EC 2.7.1.-) mek1 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 17-Mar-2000
C;Accession: T43420; T37688
R;Lyne, M.H.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z22503
A;Accession: T43420
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-445 <LYN>
A;Cross-references: EMBL:Z71478; PIDN:CAA96101.1
A;Experimental source: strain 972h(-)
R;Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z21737
A;Accession: T37688
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-445 <DEV>
A;Cross-references: EMBL:Z98530; PIDN:CAB11125.1; [illegible]; GSPD:SPAC1404.03
A;Experimental source: strain 972h-; cosmid c1404
C;Genetics:
A;Gene: mek1; SPAC1404.03
A;Map position: 1
A;Introns: [illegible]
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C;Keywords: phosphotransferase

  Query Match             74.4%;  Score 32;  DB 2;  Length 445;
  Best Local Similarity   75.0%;  Pred. No. 40;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 VQFVASYK 8
              :|| ||||
Db         75 LQFTASYK 82

RESULT 9
S15566
pol protein - simian foamy virus
C;Species: simian foamy virus
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 12-Apr-1995
C;Accession: S15566
R;Mergia, A.; Shaw, S.E.S.; Jonathan, L.E.; Paul, L.A.
submitted to the EMBL Data Library, March 1991
A;Description: Sequence relationships of simian foamy virus type 1 with human foamy v
A;Reference number: S15565
A;Accession: S15566
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-958 <MER>
A;Cross-references: EMBL:X58484
C;Superfamily: pol polyprotein

  Query Match             74.4%;  Score 32;  DB 2;  Length 958;
  Best Local Similarity   55.6%;  Pred. No. 87;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 VQFVASYKV 9
              |||:  ||:
Db        621 VQFIPEYKI 629

RESULT 10
S18738
pol protein - simian foamy virus (fragment)
C;Species: simian foamy virus
C;Date: 19-Mar-1997 #sequence_revision 24-Jul-1997 #text_change 31-Oct-1997
C;Accession: S18738
R;Kupiec, J.J.; Kay, A.; Hayat, M.; Ravier, R.; Peries, J.; Galibert, F.
Gene 101, 185-194, 1991
A;Title: Sequence analysis of the simian foamy virus type 1 genome.
A;Reference number: S18737; MUID:91276270
A;Accession: S18738
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1161 <KUP>
A;Cross-references: EMBL:X54482
C;Superfamily: pol polyprotein

  Query Match             74.4%;  Score 32;  DB 2;  Length 1161;
  Best Local Similarity   55.6%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy       1 VQFVASYKV 9
           |||:  ||:
Db     633 VQFIPEYKI 641


RESULT  11
B82260
rare lipoprotein B VC0954 [imported]  Vibrio cholerae (strain N16961 serogroup O1)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 28-Aug-2000 #text_change 02-Feb-2001
C;Accession: B82260
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, T.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M
Nature 406, 477-483, 2000
A;Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae
A;Reference number: A82245; MUID:20408033
A;Accession: B82260
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-213 <HEI>
A;Cross-references: GB:AE004177; GB:AE003852; NID:g9655398; PIDN:AAF94116.1; GSPDB:GN001
A;Experimental source: serogroup O1, strain N16961, biotype El Tor
C;Genetics:
A;Gene: VC0954
A;Map position: 1
C;Superfamily: rplB lipoprotein


  Query Match             72.1%;  Score 31;  DB 2;  Length 213;
  Best Local Similarity   85.7%;  Pred. No. 31;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 FVASYKV 9
           |||||:|
Db     113 FVASYRV 119


RESULT  12
T40730
probable rna-binding protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T40730
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lelaure, V.; Galibert, F.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z21947
A;Accession: T40730
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-217 <LYN>
A;Cross-references: EMBL:AL033388; PIDN:CAA21840.1; GSPDB:GN00867; SPDB:SPBC887.05c
A;Experimental source: strain 972h-; cosmid c887
C;Genetics:
A;Gene: SPDB:SPBC887.05c
A;Map position: 2
A;Introns: 2/2


  Query Match             72.1%;  Score 31;  DB 2;  Length 217;
  Best Local Similarity   77.8%;  Pred. No. 32;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |  |||||
Db     104 VDHVASYKV 112


RESULT  13
B82789
hypothetical protein XF0584 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 28-Aug-2000 #text_change 28-Aug-2000
C;Accession: B82789
R;Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: B82789
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-308 <SIM>
A;Cross-references: GB:AE003904; GB:AE003849; NID:g9105432; PIDN:AAF83394.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
Submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
, F.C.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0584


  Query Match             72.1%;  Score 31;  DB 2;  Length 308;
  Best Local Similarity   85.7%;  Pred. No. 46;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 FVASYKV 9
           |||||:|
Db      97 FVASYRV 103


RESULT  14
T00110
integral membrane acetyltransferase homolog  Actinobacillus actinomycetemcomitans
C;Species: Actinobacillus actinomycetemcomitans
C;Date: 22-Jun-1999 #sequence_revision 22-Jun-1999 #text_change 21-Jul-2000
C;Accession: T00110
R;Nakano, Y.; Yoshida, Y.; Yamashita, Y.; Koga, T.
Biochim. Biophys. Acta 1442, 409-414, 1998
A;Title: A gene cluster for 6-deoxy-L-talan synthesis in Actinobacillus actinomycetem
A;Reference number: Z14111; MUID:99023768
A;Accession: T00110
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-344 <NAK>
A;Cross-references: EMBL:AB010415; NID:g6142248; PIDN:BAA28139.1; PID:g3132262
A;Experimental source: strain NCTC 9710


  Query Match             72.1%;  Score 31;  DB 2;  Length 344;
  Best Local Similarity   62.5%;  Pred. No. 51;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QFVASYKV 9
           ||: :|||
Db     202 QFIVTYKV 209


RESULT  15
SYBSES
glutamate--tRNA ligase (EC 6.1.1.17) - Bacillus stearothermophilus
N;Alternate names: glutamyl-tRNA synthetase
C;Species: Bacillus stearothermophilus
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 18-Jun-1999
C;Accession: S30090
```

R;Breton, R., Watson, D., Yaguchi, M., Lapointe, J.
J. Biol. Chem. 265, 18248-18255, 1990
A;Title: Glutamyl-tRNA synthetases of Bacillus subtilis 168T and of Bacillus stearotherm
A;Reference number: A36090; MUID:91009314
A;Accession: B36090
A;Molecule type: DNA
A;Residues: 1-489 <BRE>
A;Cross-references: GB:M55072; NID:g143903; PIDN:AAA22434.1; PID:g143904; GB:J05b47
A;Note: the authors translated the codon AAC for residue 62 as Asp and CCG for residue 1
C;Genetics:
A;Gene: gltX
C;Superfamily: glutamate--tRNA ligase; glutamine--tRNA ligase homology
C;Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis
F;5-283/Domain: glutamine--tRNA ligase homology <EGL>
F;256/Binding site: ATP (Lys) #status predicted

Query Match 72.1%; Score 31; DB 1; Length 489;
Best Local Similarity 62.5%; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VQFVASYK 8
:|||: ||
Db 262 IQFVSQYK 269

Search completed: September 5, 2002, 15:28:54
Job time: 359 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:42 ; Search time 33.00 Seconds
(without alignments)
10.252 Million cell updates/sec

Title: US-09-744-804-40
Perfect score: 43
Sequence: 1 VQFVASYKV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters. 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	100.0	387	1	MFGM_HUMAN	Q08431 homo sapien
2	39	90.7	427	1	MFGM_RAT	P78490 rattus norv
3	36	83.7	409	1	MFGM_PIG	P79385 sus scrofa
4	34	79.1	463	1	MFGM_MOUSE	P21956 mus musculu
5	33	76.7	427	1	MFGM_BOVIN	Q95114 bos taurus
6	33	76.7	533	1	SYK_METMP	Q30522 methanococ
7	32	74.4	445	1	MEK1_SCHPO	Q10292 schizosacch
8	32	74.4	1161	1	POL_SFV1	P23074 simian foam
9	31	72.1	489	1	SYE_BACST	P22249 bacillus st
10	30	69.8	641	1	PRIM_UREPA	Q9ppz6 ureaplasma
11	29	67.4	145	1	DISC_DICDI	P02888 dictyosteli
12	29	67.4	223	1	HEXB_ADEP3	Q83453 porcine ade
13	29	67.4	253	1	DISA_DICDI	P02886 dictyosteli
14	29	67.4	253	1	DISC_DICDI	P02887 dictyosteli
15	29	67.4	277	1	CYST_ECOLI	P16701 escherichia
16	29	67.4	277	1	CYST_SALTY	P11032 salmonella
17	29	67.4	351	1	PML1_MOUSE	Q08790 mus musculu
18	29	67.4	436	1	ERF1_BLEAM	Q9bmm3 blepharisma
19	29	67.4	485	1	SYE_BACHD	Q9kgf6 bacillus ha
20	29	67.4	560	1	VATB_NPVOP	P11716 choristonea
21	29	67.4	560	1	VATB_NPVSP	P22114 orgyia pseu
22	29	67.4	928	1	NFP1_XENLA	P28824 xenopus lae
23	29	67.4	1170	1	ITAL_HUMAN	P20701 homo sapien
24	29	67.4	1477	1	YOR1_YEAST	P53049 saccharomyc
25	29	67.4	2194	1	POLG_HEPAT	P32537 human enter
26	28	65.1	130	1	UCR6_SCHMA	Q81674 schistosoma
27	28	65.1	159	1	OA2_CAEEL	Q9nhz6 caenorhabdi
28	28	65.1	192	1	PAS2_DROME	P04388 drosophila
29	28	65.1	204	1	PRA2_HUMAN	P17082 homo sapien
30	28	65.1	207	1	HIS5_AQUAE	O66943 aquifex aeo
31	28	65.1	218	1	PRAS_HUMAN	P10301 homo sapien
32	28	65.1	218	1	PRAS_MOUSE	P10833 mus musculu
33	28	65.1	309	1	RAS1_YEAST	P01119 saccharomyc
34	28	65.1	382	1	P4C6_HUMAN	Q9ubx8 h beta-1,4-
35	28	65.1	399	1	DNAA_AQUAE	O66659 aquifex aeo
36	28	65.1	425	1	SECY_CDOSI	P40461 odontella s
37	28	65.1	447	1	PLP_KLULA	P17283 kluyveromyc
38	28	65.1	483	1	SYE_BACSU	P22259 bacillus su
39	28	65.1	1252	1	RPOB_CHLPN	Q9z9a0 chlamydia p
40	28	65.1	2368	1	ESR1_YEAST	P38111 saccharomyc
41	28	65.1	2787	1	TEL1_YEAST	P38112 saccharomyc
42	27.5	64.0	89	1	Y08A_HAEPA	Q9pr55 ureaplasma
43	27	62.8	81	1	YW84_CAEEL	Q19327 caenorhabdi
44	27	62.8	99	1	DP30_HUMAN	Q9c005 homo sapien
45	27	62.8	99	1	DP30_MOUSE	Q991t0 mus musculu

ALIGNMENTS

RESULT 1
MFGM_HUMAN
ID MFGM_HUMAN STANDARD; PRT; 387 AA.
AC Q08431;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMFG)
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
GN MFGE8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast, and Breast carcinoma;
RX MEDLINE=96213908; PubMed=8630264;
RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
RT "Cloning and sequence analysis of human breast epithelial antigen
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal
RT growth factor-like domain.";
RL DNA Cell Biol. 15:281-286(1996).
RN [2]
RP SEQUENCE OF 170-387 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=91371351; PubMed=1909932;
RA Iaracca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
RA Ceriani R.L.;
RT "A Mr 46,000 human milk fat globule protein that is highly expressed
RT in human breast tumors contains factor VIII-like domains.";
RL Cancer Res. 51:4994-4998(1991).
RN [3]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=98144924, PubMed=9532676;
RA Giuffrida M.G., Cavaletto M., Giunta C., Conti A.,
RA Godovac-Zimmermann J.;
RT "Isolation and characterization of full and truncated forms of human
RT breast carcinoma protein BA46 from human milk fat globule membranes.";
RL J. Protein Chem. 17:143-148(1998).
RN [4]
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
RX MEDLINE=99342076, PubMed=10411933;
RA Haeggqvist B., Naslund J., Slettengren E., Westermark G.T., Mucchiano G.,
RA Tjernberg L.O., Nordstedt C., Engstrom U., Westermark P.;
RT "Medin: an integral fragment of aortic smooth muscle cell produced
RT lactadherin forms the most common human amyloid.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=97405986, PubMed=9281929,
RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
RT (RGD)-dependent cell adhesion.",

```
RL   DNA Cell Biol 16:861-869(1997).
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
CC       SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
CC   -!- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
CC       MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC   -!- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
CC       AMINO ACID 264 AND 273.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ---------------------------------------------------------------------
DR   EMBL; U58516; AAC50549.1; -.
DR   EMBL; S56151; AAB19771.1; -.
DR   MIM; 602281; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
FT   SIGNAL       1     23
FT   CHAIN       24    387       LACTADHERIN.
FT   CHAIN      202    387       LACTADHERIN, SHORT FORM.
FT   CHAIN      268    317       MEDIN.
FT   DOMAIN      24     67       EGF-LIKE.
FT   DOMAIN      70    225       F5/8 TYPE C 1.
FT   DOMAIN     230    387       F5/8 TYPE C 2.
FT   SITE        46     48       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    27     38       BY SIMILARITY.
FT   DISULFID    32     55       BY SIMILARITY.
FT   DISULFID    57     66       BY SIMILARITY.
FT   DISULFID    70    225       BY SIMILARITY.
FT   DISULFID   212    216       BY SIMILARITY.
FT   DISULFID   230    387       BY SIMILARITY.
FT   CARBOHYD   238    238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   325    325       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   329    329       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   350    350       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   387 AA;  43123 MW;  2EE6571DEC83782D CRC64;


  Query Match             100.0%;  Score 43;  DB 1;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.018;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||||||||
Db     313 VQFVASYKV 321


RESULT   2
MFGM_RAT
ID   MFGM_RAT       STANDARD;      PRT;   427 AA.
AC   P70490;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (O-
DE   acetyl GD3 ganglioside synthase) (AGS) (MFGM).
GN   MFGE8 OR AGS.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96374422; PubMed=8780713;
RA   Ogura K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
RT   "Cloning and expression of cDNA for O-acetylation of GD3
RT   ganglioside.";
RL   Biochem. Biophys. Res. Commun. 225:932-938(1996).
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO
CC       PARTICIPATE IN THE O-ACETYLATION OF GD3 GANGLIOSIDE SIALIC ACID.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; P84068; PAA12210.1; -.
DR   HSSP; P00740; 1IXA.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23    427       LACTADHERIN.
FT   DOMAIN      24     61       EGF-LIKE 1.
FT   DOMAIN      64    108       EGF-LIKE 2.
FT   DOMAIN     111    267       F5/8 TYPE C 1.
FT   DOMAIN     272    427       F5/8 TYPE C 2.
FT   DISULFID    28     39       BY SIMILARITY.
FT   DISULFID    43     49       BY SIMILARITY.
FT   DISULFID    51     60       BY SIMILARITY.
FT   DISULFID    68     79       BY SIMILARITY.
FT   DISULFID    73     96       BY SIMILARITY.
FT   DISULFID    98    107       BY SIMILARITY.
FT   DISULFID   111    267       BY SIMILARITY.
FT   DISULFID   254    258       BY SIMILARITY.
FT   DISULFID   272    427       BY SIMILARITY.
FT   SITE        87     89       CELL ATTACHMENT SITE (POTENTIAL).
FT   CARBOHYD    61     61       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   230    230       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   280    280       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   340    340       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   427 AA;  47413 MW;  EA8C863F3EE6047 CRC64;


  Query Match             90.7%;  Score 39;  DB 1;  Length 427;
  Best Local Similarity   77.8%;  Pred. No. 0.18;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           :|:|||||
Db     355 IQYVASYKV 363
```

```
RESULT   3
MFGM_PIG
ID   MFGM_PIG        STANDARD;      PRT;   409 AA.
AC   P79385;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Lactadherin (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM) (Sperm
DE   surface protein SP47) (PP47).
GN   MFGE8.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC       SPERMATOZOAN.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Y11683; CAA72379.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Glycoprotein; Repeat; EGF-like domain.
FT   DOMAIN        2     41       EGF-LIKE 1.
FT   DOMAIN       44     88       EGF-LIKE 2.
FT   DOMAIN       91    247       F5/8 TYPE C 1.
FT   DOMAIN      252    409       F5/8 TYPE C 2.
FT   SITE         67     69       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID      6     17       BY SIMILARITY.
FT   DISULFID     11     29       BY SIMILARITY.
FT   DISULFID     31     40       BY SIMILARITY.
FT   DISULFID     91    247       BY SIMILARITY.
FT   DISULFID    234    338       BY SIMILARITY.
FT   DISULFID    252    409       BY SIMILARITY.
FT   CARBOHYD     41     41       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    372    372       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   409 AA;  45725 MW;  B0C07AE860C9327A CRC64;


  Query Match             83.7%;  Score 36;  DB 1;  Length 409;
  Best Local Similarity   66.7%;  Pred. No. 0.88;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           :|:||:|||
Db     335 IQYVAAYKV 343
```

```
RESULT   4
MFGM_MOUSE
ID   MFGM_MOUSE      STANDARD;      PRT;   463 AA.
AC   P21956; P97800;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM)
DE   (Sperm surface protein SP47) (MP47).
GN   MFGE8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.
RC   TISSUE=Mammary gland;
RX   MEDLINE=91046008; PubMed=2122462;
RA   Stubbs J.D., Lekutis C., Singer K.L., Bui A., Yuzuki D.,
RA   Srinivasan U., Parry G.;
RT   "cDNA cloning of a mouse mammary epithelial cell surface protein
RT   reveals the existence of epidermal growth factor-like domains linked
RT   to factor VIII-like sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
RN   [2]
RP   SEQUENCE OF 23-463 FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC       SPERMATOZOAN.
CC   -!- DEVELOPMENTAL STAGE: MRNA EXPRESSION IS DETECTABLE IN MAMMARY
CC       TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M38337; AAA29534.1; -.
DR   EMBL; Y11684; CAA72380.1; -.
DR   PIR; A36479; A36479.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:102768; Mfge8.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
FT   SIGNAL        1     22
FT   CHAIN        23    463       LACTADHERIN.
FT   DOMAIN       24     61       EGF-LIKE 1.
FT   DOMAIN       64    108       EGF-LIKE 2.
FT   DOMAIN      148    303       F5/8 TYPE C 1.
FT   DOMAIN      308    463       F5/8 TYPE C 2.
```

```
FT   SITE          87     89       CELL ATTACHMENT SITE (POTENTIAL)
FT   DISULFID      28     39       BY SIMILARITY.
FT   DISULFID      33     49       BY SIMILARITY.
FT   DISULFID      51     60       BY SIMILARITY.
FT   DISULFID      68     79       BY SIMILARITY.
FT   DISULFID      73     96       BY SIMILARITY.
FT   DISULFID      98    107       BY SIMILARITY.
FT   DISULFID     148    303       BY SIMILARITY.
FT   DISULFID     290    294       BY SIMILARITY.
FT   DISULFID     308    463       BY SIMILARITY.
FT   CARBOHYD      61     61       N-LINKED (GLCNAC ...) (POTENTIAL).
FT   CARBOHYD     266    266       N-LINKED (GLCNAC ...) (POTENTIAL).
FT   CARBOHYD     316    316       N-LINKED (GLCNAC ...) (POTENTIAL).
FT   CARBOHYD     426    426       N-LINKED (GLCNAC ...) (POTENTIAL).
FT   CONFLICT      30     30       S -> F (IN REF. 2).
FT   CONFLICT      35     35       N -> D (IN REF. 1; AA SEQUENCE).
FT   CONFLICT     110    147       ETNYYNLDGEYMFTTAVPNTAVPTPAPTPDLSNNLASR ->
FT                                 G (IN REF. 2).
FT   CONFLICT     168    168       Y -> S (IN REF. 2).
FT   CONFLICT     196    196       H -> P (IN REF. 2).
FT   CONFLICT     309    309       L -> S (IN REF. 2).
FT   CONFLICT     395    395       E -> A (IN REF. 2).
SQ   SEQUENCE   463 AA;  51465 MW;  D78B6C6EFBBA724D CRC64;


  Query Match             79.1%;  Score 34;  DB 1;  Length 463;
  Best Local Similarity   66.7%;  Pred. No. 3;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           :|:| ||||
Db     391 IQYVESYKV 399


RESULT   5
MFGM_BOVIN
ID   MFGM_BOVIN     STANDARD;      PRT;   427 AA.
AC   Q95114; Q27959; P79344;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-OCT-2001 (Rel. 40, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
DE   (MGP57/53) (PAS-6/PAS-7 glycoprotein) (MPCM) (Sperm surface protein
DE   SP47) (BP47) (Components 15/16).
GN   MFGE8.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC   STRAIN=HOLSTEIN; TISSUE=Mammary gland;
RX   MEDLINE=97008954; PubMed=8856064;
RA   Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RA   Petersen T.E.;
PT   "Characterization of glycoprotein PAS-6/7 from membranes of bovine
PT   milk fat globules.";
RL   Eur. J. Biochem. 240:628-636(1996).
RN   [2]
RP   SEQUENCE OF 18-427 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=96125736; PubMed=8541316;
RA   Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA   Matsuda T.;
PT   "Molecular cloning of glycoprotein antigens MGP57/53 recognized by
PT   monoclonal antibodies raised against bovine milk fat globule
PT   membrane.";
RL   Biochim. Biophys. Acta 1245:385-391(1995).
RN   [3]
RP   SEQUENCE OF 19-427 FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
RC   TISSUE=Milk;
RX   MEDLINE=93258576; PubMed=8485470;
RA   Mather I.H., Banghart L.R., Lane W.S.;
PT   "The major fat-globule membrane proteins, bovine components 15/16 and
PT   guinea pig GP 55, are homologous to MGF-E8, a murine glycoprotein
PT   containing epidermal growth factor-like and factor V/VIII like
PT   sequences.";
RL   Biochem. Mol. Biol. Int. 29:545-554(1993).
CC   -!- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
CC       MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM
CC       LACKS 53 AMINO ACIDS WITHIN THE F5/8 TYPE C 1 DOMAIN.
CC   -!- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
CC   -!- PTM: THE 2 O-LINKED GLYCANS CONSIST OF GAL, GLCNAC AND FUC, WITH
CC       PROBABLY FUC AS REDUCING TERMINAL SUGAR.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X91895; CAA62997.1; -.
DR   EMBL; S80643; AAB35894.2; -.
DR   EMBL; Y11719; CAA72406.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain;
KW   Alternative splicing.
FT   SIGNAL        1     18
FT   CHAIN        19    427       LACTADHERIN.
FT   DOMAIN       20     59       EGF-LIKE 1.
FT   DOMAIN       62    106       EGF-LIKE 2.
FT   DOMAIN      109    265       F5/8 TYPE C 1.
FT   DOMAIN      270    427       F5/8 TYPE C 2.
FT   SITE         85     87       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     24     35       BY SIMILARITY.
FT   DISULFID     29     47       BY SIMILARITY.
FT   DISULFID     49     58       BY SIMILARITY.
FT   DISULFID     66     77       BY SIMILARITY.
FT   DISULFID     71     94       BY SIMILARITY.
FT   DISULFID     96    105       BY SIMILARITY.
FT   DISULFID    109    265
FT   DISULFID    252    256
FT   DISULFID    270    427
FT   CARBOHYD     27     27       O-LINKED (FUC. ...) (IN PAS-6).
FT   CARBOHYD     34     34       O-LINKED (FUC. ...) (IN PAS-7).
FT   CARBOHYD     53     53       N-LINKED (GLCNAC ...) (HYBRID) (IN PAS-6
FT                                AND PAS-7).
FT   CARBOHYD    227    227       N-LINKED (GLCNAC ...) (HIGH MANNOSE) (IN
FT                                PAS-6).
FT   VARSPLIC    169    221       MISSING (IN SHORT ISOFORM).
FT   CONFLICT     19     19       A -> F (IN REF. 1).
FT   CONFLICT     28     28       I -> Q (IN REF. 1).
SQ   SEQUENCE   427 AA;  47411 MW;  4CBBEE3A1DC4EB24 CRC64;
```

```
  Query Match             76.7%;  Score 33;  DB 1;  Length 427;
  Best Local Similarity   55.6%;  Pred. No. 4.7;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           :|:|||:|:|
Db     353 IQYVAAYRV 361


RESULT   6
SYK_METMP
ID   SYK_METMP      STANDARD;      PRT;   533 AA.
AC   O30522;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Lysyl tRNA synthetase (EC 6.1.1.6) (Lysine tRNA ligase) (LYSRS).
GN   LYSS.
OS   Methanococcus maripaludis.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=39152;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-22.
RC   STRAIN=JJ;
RX   MEDLINE=98016292; PubMed=9384292;
RA   Ibba M., Morgan S., Curnow A.W., Pridmore D.R., Vothknecht U.C.,
RA   Gardner W., Lin W., Woese C.R., Soll D.;
RT   "A euryarchaeal lysyl-tRNA synthetase: resemblance to class I
RT   synthetases.";
RL   Science 278:1119-1122(1997).
CC   -!- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate
CC       + L-lysyl-tRNA(Lys).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF009824; AAB87410.1; -.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002904; tRNA-synt_lys_1.
DR   Pfam; PF01921; tRNA-synt_1i; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT   SITE        28     29       "HIGH" REGION.
FT   SITE       278    282       "KMSKS" REGION.
FT   CONFLICT    11     11       I -> L (IN REF. 1; AA SEQUENCE).
SQ   SEQUENCE   533 AA;  61273 MW;  807FA86BA57A0AC2 CRC64;


  Query Match             76.7%;  Score 33;  DB 1;  Length 533;
  Best Local Similarity   85.7%;  Pred. No. 6;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 FVASYKV 9
           |||||:
Db     497 FVASYKI 503


RESULT   7
MEK1_SCHPO
ID   MEK1_SCHPO     STANDARD;      PRT;   445 AA.
AC   Q10292;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Meiosis specific serine/threonine protein kinase mek1 (EC 2.7.1.37).
GN   MEK1 OR SPAC14C4.03.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Lyne M.H., Bryant J.A., Aves S.J.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Doolin K., Chatcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROBABLE PROTEIN KINASE REQUIRED FOR MEIOTIC
CC       RECOMBINATION.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   -!- SIMILARITY: CONTAINS 1 FHA DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z71478; CAA96101.1; -.
DR   EMBL; Z98596; CAB11196.1; -.
DR   HSSP; P00518; 1PHK.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR000253; FHA_domain.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
KW   Transferase; Serine/threonine protein kinase; ATP-binding; Meiosis.
FT   DOMAIN      62    116       FHA.
FT   DOMAIN     160    421       PROTEIN KINASE.
FT   NP_BIND    166    174       ATP (BY SIMILARITY).
FT   ACT_SITE   281    281       BY SIMILARITY.
FT   BINDING    189    189       ATP (BY SIMILARITY).
SQ   SEQUENCE   445 AA;  51151 MW;  6DE2D795DA1B45B0 CRC64;


  Query Match             74.4%;  Score 32;  DB 1;  Length 445;
  Best Local Similarity   75.0%;  Pred. No. 8.5;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYK 8
           :|| ||||
Db      75 IQFTASYK 82


RESULT   8
POL_SFV1
ID   POL_SFV1       STANDARD;      PRT;  1161 AA.
AC   P23074;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Pol polyprotein [Contains: Protease (EC 3.4.23.-); Reverse
DE   transcriptase (EC 2.7.7.49); Endonuclease].
```

```
GN   POL.
OS   Simian foamy virus (type 1) (SFV-1).
OC   Viruses; Retroid viruses; Retroviridae; Spumavirus
OX   NCBI_TaxID=11964;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91276270; PubMed=1647358;
RA   Kupiec J.-J., Kay A., Hayat M., Ravier R., Peries J., Galibert F.;
RT   "Sequence analysis of the simian foamy virus type 1 genome.";
RL   Gene 101:185-194(1991).
RN   [2]
RP   SEQUENCE OF 1-970 FROM N.A.
RX   MEDLINE=91361538; PubMed=1653483;
RA   Mergia A., Luciw P.A.;
RT   "Replication and regulation of primate foamy viruses.";
RL   Virology 184:475-482(1991).
RN   [3]
RP   SEQUENCE OF 969-1161 FROM N.A.
RX   MEDLINE=90080148; PubMed=2152825;
RA   Mergia A., Shaw K.E.S., Lackner J.E., Luciw P.A.;
RT   "Relationship of the env genes and the endonuclease domain of the pol
RT   genes of simian foamy virus type 1 and human foamy virus.";
RL   J. Virol. 64:406-410(1990).
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A9.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X54482; -; NOT_ANNOTATED_CDS.
DR   EMBL; X58484; CAA41394.1; -.
DR   EMBL; M33561; AAA47793.1; -.
DR   PIR; A33562; A33562.
DR   PIR; S15566; S15566.
DR   PIR; S18738; S18738.
DR   HSSP; P03355; 1MML.
DR   MEROPS; A09.001; -.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR000477; RVT.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR001641; Spuma_A9PTase.
DR   Pfam; PF00075; rnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00078; rvt; 1.
DR   PRINTS; PR00920; SPUMVIRPTASE.
KW   Transferase; RNA directed DNA polymerase; Hydrolase; Nuclease;
KW   Endonuclease; Aspartyl protease; Polyprotein.
FT   ACT_SITE     36     36       BY SIMILARITY.
FT   CONFLICT    236    236       T -> I (IN REF. 2).
FT   CONFLICT    910    910       S -> G (IN REF. 2).
FT   CONFLICT    950    950       A -> T (IN REF. 2).
SQ   SEQUENCE   1161 AA;  131717 MW;  08D09D1D6BB4C14C CRC64;

  Query Match             74.4%;  Score 32;  DB 1;  Length 1161;
  Best Local Similarity   55.6%;  Pred. No. 25;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 VQFVASYKV 9
            |||:  ||:
Db      633 VQFIPEYKI 641


RESULT 9
SYE_BACST
ID   SYE_BACST      STANDARD;      PRT;   489 AA.
AC   P22249;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)
DE   (GluRS).
GN   GLTX.
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Geobacillus.
OX   NCBI_TaxID=1422;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91009314; PubMed=2120226;
RA   Breton R., Watson D., Yaguchi M., Lapointe J.;
RT   "Glutamyl-tRNA synthetases of Bacillus subtilis 168T and of Bacillus
RT   stearothermophilus. Cloning and sequencing of the gltX genes and
RT   comparison with other aminoacyl tRNA synthetases.";
RL   J. Biol. Chem. 265:18248-18255(1990).
CC   -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC       diphosphate + L-glutamyl-tRNA(Glu).
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M55073; AAA22494.1; -.
DR   PIR; B36090; SYBSES.
DR   HSSP; P27000; 1GLN.
DR   InterPro; IPR000924; tRNA-synt_1c.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00749; tRNA-synt_1c; 1.
DR   PRINTS; PR00987; TRNASYNTHGLU.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT   SITE         11     21       "HIGH" REGION.
FT   SITE        253    257       "KMSKS" REGION.
FT   BINDING     256    256       ATP (BY SIMILARITY).
SQ   SEQUENCE   489 AA;  56183 MW;  AB71266CB8EA2F15 CRC64;

  Query Match             72.1%;  Score 31;  DB 1;  Length 489;
  Best Local Similarity   62.5%;  Pred. No. 16;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 VQFVASYK 8
            :|||. ||
Db      262 IQFVSQYK 269


RESULT 10
PRIM_UREPA
ID   PRIM_UREPA     STANDARD;      PRT;   641 AA.
AC   Q9PPZ6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA primase (EC 2.7.7.-).
GN   DNAG OR UU494.
OS   Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Ureaplasma.
OX   NCBI_TaxID=134821;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SEROVAR 3;
RX   MEDLINE=20500219; PubMed=11048724;
RA   Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
```

```
RA   Cassell G.H.;
RT   "The complete sequence of the mucosal pathogen Ureaplasma
RT   urealyticum.";
RL   Nature 407:757-762(2000).
CC   -!- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL
CC       RNA PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT
CC       REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.
CC   -!- COFACTOR: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   --------------------------------------------------------------------------
DR   EMBL; AE002147; AAF30906.1; -.
DR   HSSP; Q9X4D0; 1D0Q.
DR   InterPro; IPR002936; Toprim.
DR   InterPro; IPR002694; ZnF_CHCC.
DR   Pfam; PF01751; Toprim; 1.
DR   Pfam; PF01807; zf-CHC2; 1.
DR   ProDom; PD002988; ZnF_CHCC; 1.
DR   SMART; SM00493; TOPRIM; 1.
DR   SMART; SM00400; ZnF_CHCC; 1
KW   Transferase; DNA replication; DNA-directed RNA polymerase; Primosome;
KW   Zinc finger; Zinc; Metal binding; Complete proteome.
FT   ZN_FING      39     63       CHC2-TYPE (BY SIMILARITY).
SQ   SEQUENCE   641 AA;  75390 MW;  A6920168DB58EF47 CRC64;

  Query Match             69.8%;  Score 30;  DB 1;  Length 641;
  Best Local Similarity   62.5%;  Pred. No. 38;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 VQFVASYK 8
                ::||| ||
Db         70 IKFVAEYK 77

RESULT 11
DISD_DICDI
ID   DISD_DICDI     STANDARD;      PRT;   149 AA.
AC   P02888;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Discoidin I, D chain (Fragment).
GN   DSCD.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82170475; PubMed=6279874;
RA   Poole S., Firtel R.A., Lamar E., Rowekamp W.;
RT   "Sequence and expression of the discoidin I gene family in
RT   Dictyostelium discoideum.";
RL   J. Mol. Biol. 153:273-289(1981).
RN   [2]
RP   CELL ATTACHMENT SITE.
RX   MEDLINE=85074463; PubMed=6509552;
RA   Springer W.R., Cooper D.N.W., Barondes S.H.;
RT   "Discoidin I is implicated in cell-substratum attachment and ordered
RT   cell migration of Dictyostelium discoideum and resembles
RT   fibronectin.";
RL   Cell 39:557-564(1984).
CC   -!- FUNCTION: GALACTOSE- AND N-ACETYLGALACTOSAMINE-BINDING LECTIN. MAY
CC       PLAY A ROLE IN CELL-SUBSTRATUM ADHESION RATHER THAN IN CELL-CELL
CC       ADHESION. MAY BE NECESSARY FOR THE MAINTENANCE OF NORMAL ELONGATE
CC       MORPHOLOGY DURING AGGREGATION.
CC   -!- SUBUNIT: TETRAMER OF FOUR DIFFERENT CHAINS (A TO D).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: STALK CELLS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J01285; AAA33200.1; -.
DR   PIR; A03383; DLDOID.
DR   DictyDb; DD02062; dscD.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00754; F5_F8_type_C; 1.
DR   SMART; SM00231; FA58C; 1.
DR   PROSITE; PS01285; FA58C_1; 1.
DR   PROSITE; PS01286; FA58C_2; FALSE_NEG.
KW   Lectin; Cell adhesion.
FT   SITE         79     81       CELL ATTACHMENT SITE.
FT   NON_TER     149    149
SQ   SEQUENCE   149 AA;  16746 MW;  1B8E0F220B29E9A1 CRC64;

  Query Match             67.4%;  Score 29;  DB 1;  Length 149;
  Best Local Similarity   62.5%;  Pred. No. 13;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 QFVASYKV 9
                |:| |||:
Db         84 QWVTSYKI 91

RESULT 12
HEX8_ADEP3
ID   HEX8_ADEP3     STANDARD;      PRT;   223 AA.
AC   Q83453; Q9YTR6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hexon-associated protein precursor (Protein VIII).
GN   PVIII.
OS   Porcine adenovirus type 3 (PAV-3).
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX   NCBI_TaxID=35265;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6618;
RX   MEDLINE=95351046; PubMed=7625129;
RA   Reddy P.S., Nagy E., Derbyshire J.B.;
RT   "Sequence analysis of putative pVIII, E3 and fibre regions of porcine
RT   adenovirus type 3.";
RL   Virus Res. 36:97-106(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6618;
RX   MEDLINE=99058141; PubMed=9837805;
RA   Reddy P.S., Idamakanti N., Song J.Y., Lee J.B., Hyun B.H., Park J.H.,
RA   Cha S.H., Bae Y.T., Tikoo S.K., Babiuk L.A.;
RT   "Nucleotide sequence and transcription map of porcine adenovirus type
RT   3.";
RL   Virology 251:414-426(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6618 / IAF;
RA   Larocque D., Malenfant F., Massie B., Dea S.;
RT   "Porcine adenovirus serotype 3, complete genome.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ------------------------------------------------------------------------
DR   EMBL; AF083132; AAC99446.1; -.
DR   EMBL; AJ237815; CAB41034.1; -.
DR   EMBL; AB026117; BAA76972.1; -.
DR   InterPro; IPR000646; Adenc_PVIII.
DR   Pfam; PF01310; Adeno_PVIII; 1.
KW   Hexon-associated protein.
FT   PROPEP          1    111       BY SIMILARITY.
FT   CHAIN         112    223       HEXON ASSOCIATED PROTEIN.
FT   SITE          111    112       CLEAVAGE (BY ADENOVIRUS PROTEASE).
SQ   SEQUENCE   223 AA;  24281 MW;  7781342B115602A1 CRC64;


  Query Match             67.4%;  Score 29;  DB 1;  Length 223;
  Best Local Similarity   50.0%;  Pred. No. 29;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QFVASYKV 9
            ||:|:| :
Db     207 QFIANYNI 214


RESULT  13
DISA_DICDI
ID   DISA_DICDI     STANDARD;      PRT;   253 AA.
AC   P02886;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Discoidin I, A chain.
GN   DSCA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82170475; PubMed=6279874;
RA   Poole S., Firtel R.A., Lamar E., Rowekamp W.;
RT   "Sequence and expression of the discoidin I gene family in
RT   Dictyostelium discoideum.";
RL   J. Mol. Biol. 153:273-289(1981).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RX   MEDLINE=83059644; PubMed=6754951;
RA   Jellinghaus U., Schaetzle U., Schmid W., Rowekamp W.;
RT   "Transcription of a dictyostelium discoidin-i gene in yeast
RT   alternative promoter sites used in two different eukaryotic cells.";
RL   J. Mol. Biol. 159:623-636(1982).
RN   [3]
RP   SEQUENCE OF 1-40 FROM N.A.
RX   MEDLINE=92233698; PubMed=6284373;
RA   Devine J.M., Tsang A.S., Williams J.G.;
RT   "Differential expression of the members of the discoidin I multigene
RT   family during growth and development of Dictyostelium discoideum.";
RL   Cell 28:793-800(1982).
RN   [4]
RP   CELL ATTACHMENT SITE.
RX   MEDLINE=85074463; PubMed=6509552.
RA   Springer W.R., Cooper D.N.W., Barondes S.H.;
RT   "Discoidin I is implicated in cell-substratum attachment and ordered
RT   cell migration of Dictyostelium discoideum and resembles
RT   fibronectin.";
RL   Cell 39:557-564(1984).
CC   -!- FUNCTION: GALACTOSE- AND N-ACETYLGALACTOSAMINE-BINDING LECTIN. MAY
CC       PLAY A ROLE IN CELL SUBSTRATUM ADHESION RATHER THAN IN CELL-CELL
CC       ADHESION. MAY BE NECESSARY FOR THE MAINTENANCE OF NORMAL ELONGATE
CC       MORPHOLOGY DURING AGGREGATION.
CC   -!- SUBUNIT: TETRAMER OF FOUR DIFFERENT CHAINS (A TO D).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: STALK CELLS.
CC   -!- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; J01282; AAA33197.1; -.
DR   PIR; A03381; DLDO1A.
DR   DictyDb; DP02000; dscA.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00754; F5_F8_type_C; 1.
DR   SMART; SM00231; FA58C; 1.
DR   PROSITE; PS01285; FA58C_1; 1.
DR   PROSITE; PS01286; FA58C_2; FALSE_NEG.
KW   Lectin; Cell adhesion.
FT   DOMAIN          2    151       F5/8 TYPE C.
FT   SITE           79     81       CELL ATTACHMENT SITE.
SQ   SEQUENCE   253 AA;  28239 MW;  81263A100F1EE91E CRC64;


  Query Match             67.4%;  Score 29;  DB 1;  Length 253;
  Best Local Similarity   62.5%;  Pred. No. 24;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QFVASYKV 9
            |:| |||:
Db      84 QWVTSYKI 91


RESULT  14
DISC_DICDI
ID   DISC_DICDI     STANDARD;      PRT;   253 AA.
AC   P02887;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Discoidin I, C chain and B chain.
GN   DSCB AND DSCC.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A. (C CHAIN AND B CHAIN 1-149).
RX   MEDLINE=82170475; PubMed=6279874;
RA   Poole S., Firtel R.A., Lamar E., Rowekamp W.;
RT   "Sequence and expression of the discoidin I gene family in
RT   Dictyostelium discoideum.";
RL   J. Mol. Biol. 153:273-289(1981).
RN   [2]
RP   CELL ATTACHMENT SITE.
RX   MEDLINE=85074463; PubMed=6509552;
RA   Springer W.R., Cooper D.N.W., Barondes S.H.;
RT   "Discoidin I is implicated in cell substratum attachment and ordered
RT   cell migration of Dictyostelium discoideum and resembles
RT   fibronectin.";
RL   Cell 39:557-564(1984).
CC   -!- FUNCTION: GALACTOSE- AND N-ACETYLGALACTOSAMINE BINDING LECTIN. MAY
CC       PLAY A ROLE IN CELL-SUBSTRATUM ADHESION RATHER THAN IN CELL-CELL
CC       ADHESION. MAY BE NECESSARY FOR THE MAINTENANCE OF NORMAL ELONGATE
CC       MORPHOLOGY DURING AGGREGATION.
CC   -!- SUBUNIT: TETRAMER OF FOUR DIFFERENT CHAINS (A TO D).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: STALK CELLS.
CC   -!- SIMILARITY: THE SEQUENCE OF RESIDUES 1-149 OF THE B CHAIN IS
CC       IDENTICAL WITH THAT OF THE C CHAIN.
CC   -!- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
CC   ------------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ------------------------------------------------------------------------
DR   EMBL; J01284; AAA33199.1; -.
DR   EMBL; J01283; AAA33198.1; -.
DR   PIR; A03382; DLDOIC.
DR   PIR; B03382; B03382.
DR   SWISS-2DPAGE; P02887. DICTY.
DR   DictyDb; DD02001; dscC.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00754; F5_F8_type_C; 1.
DR   SMART; SM00231; FA58C; 1.
DR   PROSITE; PS01285; FA58C_1; 1.
DR   PROSITE; PS01286; FA58C_2; FALSE_NEG
KW   Lectin; Cell adhesion.
FT   DOMAIN        2    151       F5/8 TYPE C
FT   SITE         79     81       CELL ATTACHMENT SITE.
SQ   SEQUENCE   253 AA;  28392 MW;  875EA9481F85444 CRC64;


  Query Match             67.4%;  Score 29;  DB 1;  Length 253;
  Best Local Similarity   62.5%;  Pred. No. 23;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QFVASYKV 9
            |:| |||:
Db      84 QWVTSYKI 91


RESULT  15
CYST_ECOLI
ID   CYST_ECOLI     STANDARD;      PRT;   277 AA.
AC   P16701;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Sulfate transport system permease protein cysT.
GN   CYSU OR CYST OR B2424
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=90264334; PubMed=2188958;
RA   Sirko A., Hryniewicz M.M., Hulanicka D.M., Bolen A.;
RT   "Sulfate and thiosulfate transport in Escherichia coli K-12:
RT   nucleotide sequence and expression of the cysTWAM gene cluster.";
RL   J. Bacteriol. 172:3351-3357(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97349980; PubMed=9205837;
RA   Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA   Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA   Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA   Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA   Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA   Yamagata S., Horiuchi T.;
RT   "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT   -K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT   analysis of its sequence features.",
RL   DNA Res. 4:91-113(1997).
CC   -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC       FOR SULFATE AND THIOSULFATE. PROBABLY RESPONSIBLE FOR THE
CC       TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
CC       PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE CYSTW
CC       SUBFAMILY.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; M32101; AAA23637.1; -.
DR   EMBL; AE000330; AAC75477.1; -.
DR   EMBL; D90871; BAA16298.1; -.
DR   EMBL; D90872; BAA16307.1; -.
DR   PIR; A35402; QRECST.
DR   PIR; B35403; B35403.
DR   EcoGene; EG10747; cysU.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF00528; BPD_transp; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
KW   Inner membrane; Transmembrane; Sulfate transport; Transport;
KW   Complete proteome.
FT   TRANSMEM     17     37       POTENTIAL.
FT   TRANSMEM     64     84       POTENTIAL.
FT   TRANSMEM     99    119       POTENTIAL.
FT   TRANSMEM    136    156       POTENTIAL.
FT   TRANSMEM    185    205       POTENTIAL.
FT   TRANSMEM    215    235       POTENTIAL.
FT   TRANSMEM    243    263       POTENTIAL.
SQ   SEQUENCE   277 AA;  30231 MW;  139282B0DE24459 CRC64;


  Query Match             67.4%;  Score 29;  DB 1;  Length 277;
  Best Local Similarity   75.0%;  Pred. No. 26;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 QFVASYKV 9
            | ||:|||
Db      55 QVVAAYKV 62


Search completed: September  5, 2002, 15:31:43
Job time: 478 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:00 ; Search time 122.98 seconds
(without alignments)
12.673 Million cell updates/sec

Title: US-09-744-804-40
Perfect score: 43
Sequence: 1 VQFVASYKV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172991929 residues

Total number of hits satisfying chosen parameters. 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39	90.7	426	11	Q9WTS3	Q9wts3 mus musculu
2	39	90.7	463	11	Q9R1X9	Q9r1x9 mus musculu
3	36	83.7	480	4	O43854	O43854 homo sapien
4	36	83.7	480	11	O35474	O35474 mus musculu
5	32	74.4	330	5	Q960N3	Q960n3 drosophila
6	32	74.4	483	5	Q9VMA1	Q9vma1 drosophila
7	32	74.4	610	5	Q9VC81	Q9vc81 drosophila
8	31	72.1	121	12	Q91932	Q91932 culex nigri
9	31	72.1	213	16	Q9KTE8	Q9kte8 vibrio chol
10	31	72.1	217	3	O94290	O94290 schizosacch
11	31	72.1	308	16	Q9PFS3	Q9pfs3 xylella fas
12	31	72.1	310	4	Q9NWI9	Q9nwi9 homo sapien
13	31	72.1	344	2	O66257	O66257 actinobacil
14	31	72.1	1206	11	Q91YJ6	Q91yj6 mus musculu
15	31	72.1	1226	17	Q58836	Q58836 methanococc
16	30	69.8	134	5	Q9VKQ9	Q9vkq9 drosophila
17	30	69.8	159	10	Q94BL0	Q94bl0 vitis vinif
18	30	69.8	193	9	Q9MCE8	Q9mce8 bacteriopha
19	30	69.8	260	16	Q9WQZ4	Q9wqz4 bacillus ha
20	30	69.8	290	10	Q9ZR10	Q9zr10 arabidopsis
21	30	69.8	296	10	Q39099	Q39099 arabidopsis
22	30	69.8	358	10	Q9FYX1	Q9fyx1 lycopersico
23	30	69.8	363	6	O77718	O77718 equus caba
24	30	69.8	374	16	Q9PB50	Q9pb50 xylella fas
25	30	69.8	386	17	Q9V169	Q9v169 pyrococcus
26	30	69.8	406	2	Q9F797	Q9f797 salmonella
27	30	69.8	476	4	Q9ULZ4	Q9ulz4 homo sapien
28	30	69.8	476	4	Q9UJW2	Q9ujw2 homo sapien
29	30	69.8	485	2	Q93171	Q93171 escherichia
30	30	69.8	529	5	Q18748	Q18748 caenorhabdi
31	30	69.8	576	5	P91296	P91296 caenorhabdi
32	30	69.8	672	5	Q02164	Q02164 caenorhabdi
33	30	69.8	967	5	Q77305	Q77305 plasmodium
34	30	69.8	1216	3	O59901	O59901 emericella
35	30	69.8	2162	5	Q9W2V6	Q9w2v6 drosophila
36	29	67.4	97	6	Q9N163	Q9n163 papio hamad
37	29	67.4	106	12	Q98499	Q98499 paramecium
38	29	67.4	130	10	Q9FKM0	Q9fkm0 arabidopsis
39	29	67.4	131	8	Q9TA16	Q9ta16 cafeteria r
40	29	67.4	214	8	Q9MLI1	Q9mli1 paraqonimus
41	29	67.4	222	12	Q89735	Q89735 sigma virus
42	29	67.4	223	12	Q64870	Q64870 porcine ade
43	29	67.4	229	2	Q53710	Q53710 staphyloco
44	29	67.4	274	5	Q965R8	Q965r8 caenorhabdi
45	29	67.4	276	2	O30833	O30833 rhodobacter

ALIGNMENTS

RESULT 1
Q9WTS3
ID Q9WTS3 PRELIMINARY; PRT; 426 AA.
AC Q9WTS3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MILK FAT GLOBULE GLYCOPROTEIN MFG-E8 S (SIMILAR TO MILK FAT GLOBULE-
DE EGF FACTOR 8 PROTEIN).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=MAMMARY GLAND;
RX MEDLINE=99129894; PubMed=9920772;
RA Oshima K., Aoki N., Negi M., Kishi M., Kitajima K., Matsuda T.;
RT "Lactation-dependent expression of an mRNA splice variant with an exon
RT for a multiply O-glycosylated domain of mouse milk fat globule
RT glycoprotein MFG-E8.";
RL Biochem. Biophys. Res. Commun. 254:522-528(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; AB025280; BAA76386.1; -.
DR EMBL; BC003904; AAH03904.1; -.
DR EMBL; BC003892; AAH03892.1; -.
DR HSSP; P00740; 1EDM.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR001092; HLH_dim.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR PRINTS; PR00010; EGFBLOOD.
DR SMART; SM00181; EGF; 2.

```
DR   SMART; SM00231; FA58C; 2
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   EGF-like domain; Glycoprotein
SQ   SEQUENCE   426 AA;  47157 MW;  E282F3A2962981A CRC64;

  Query Match             90.7%;  Score 39;  DB 11;  Length 426;
  Best Local Similarity   77.8%;  Pred. No. 2.8;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 VQFVASYKV 9
            :|:||||||
Db      354 IQYVASYKV 362

RESULT  2
Q9R1X9
ID   Q9R1X9      PRELIMINARY;      PRT;   463 AA.
AC   Q9R1X9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MILK FAT GLOBULE GLYCOPROTEIN MFG-E8 LONG FORM
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN BALB/C; TISSUE=MAMMARY GLAND;
RX   MEDLINE=99120894; PubMed 9920772;
RA   Oshima K., Aoki N., Negi M., Kishi M., Kitajima K., Matsuda T.;
RT   "Lactation-dependent expression of an mRNA splice variant with an exon
RT   for a multiply O-glycosylated domain of mouse milk fat globule
RT   glycoprotein MFG-E8.";
RL   Biochem. Biophys. Res. Commun. 254:522-528(1999).
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS
DR   EMBL; AB021130; BAA35180.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR001092; HLH_dim.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00181; EGF; 2
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; 2
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   EGF-like domain; Glycoprotein
SQ   SEQUENCE   463 AA;  51269 MW;  D71902BE09EE6427 CRC64;

  Query Match             90.7%;  Score 39;  DB 11;  Length 463;
  Best Local Similarity   77.8%;  Pred. No. 3.1;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 VQFVASYKV 9
            :|:||||||
Db      391 IQYVASYKV 399

RESULT  3
O43854
ID   O43854      PRELIMINARY;      PRT;   480 AA.
AC   O43854; O43855;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   INTEGRIN-BINDING PROTEIN DEL1 PRECURSOR.
GN   DEL1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RC   TISSUE=EMBRYONIC LUNG;
RX   MEDLINE 98083109; PubMed 9420328;
RA   Hidai C., Zupancic T.J., Penta K., Mikhail A., Kawana M.,
RA   Quertermous E.E., Aoka Y., Fukagawa M., Matsui Y., Platika D.,
RA   Auerbach R., Hogan B.L.M., Snodgrass R., Quertermous T.;
RT   "Cloning and characterization of developmental endothelial locus-1: an
RT   embryonic endothelial cell protein that binds the alphavbeta3 integrin
RT   receptor.";
RL   Genes Dev. 12:21-33(1998).
CC   -!- FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH
CC       INTERACTION WITH THE ALPHA-V BETA3 INTEGRIN RECEPTOR. INHIBITS
CC       FORMATION OF VASCULAR-LIKE STRUCTURES. MAY BE INVOLVED IN
CC       REGULATION OF VASCULAR MORPHOGENESIS OF REMODELING IN EMBRYONIC
CC       DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX
CC   -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS; LONG ISOFORM (SHOWN HERE) AND
CC       SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR   EMBL; U70312; AAC02648.1; -.
DR   EMBL; U70313; AAC02649.1; -.
DR   HSSP; P12259; 1CZT.
DR   InterPro; IPR000152; Asx_hydroxyl
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00754; F5_F8_type_C; 2
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00001; EGF_like; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_2.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   EGF-like domain; Alternative splicing; Signal; Developmental protein;
KW   Cell adhesion; Repeat; Vascularization
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    480       INTEGRIN-BINDING PROTEIN DEL1.
FT   DOMAIN       26     59       EGF-LIKE 1
FT   DOMAIN       78    116       EGF-LIKE 2.
FT   DOMAIN      123    154       EGF-LIKE 3.
FT   DOMAIN      161    311       F5/8 TYPE C 1
FT   DOMAIN      322    473       F5/8 TYPE C 2
FT   SITE         96     98       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     26     37       BY SIMILARITY.
FT   DISULFID     31     48       BY SIMILARITY.
FT   DISULFID     50     59       BY SIMILARITY
FT   DISULFID     78     89       BY SIMILARITY.
FT   DISULFID     83    105       BY SIMILARITY.
FT   DISULFID    107    116       BY SIMILARITY.
FT   DISULFID    158    314       BY SIMILARITY.
FT   DISULFID    301    305       BY SIMILARITY
FT   DISULFID    319    476       BY SIMILARITY.
FT   VARSPLIC     66     66       A -> G (IN SHORT ISOFORM).
FT   VARSPLIC     67     76       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   480 AA;  53765 MW;  F7171E23A309FD48 CRC64;

  Query Match             83.7%;  Score 36;  DB 4;  Length 480;
```

```
  Best Local Similarity   77.8%;  Pred. No. 15;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0.

Qy       1 VQFVASYKV 9
           |||| |||:
Db     402 VQFVGSYKL 410


RESULT   4
O35474
ID   O35474      PRELIMINARY;      PRT;   480 AA.
AC   O35474; O35475;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   INTEGRIN BINDING PROTEIN DEL1 PRECURSOR.
GN   EDIL3 OR DEL1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RC   TISSUE=EMBRYO;
RX   MEDLINE=98083109; PubMed=9420328;
RA   Hidai C., Zupancic T., Penta K., Mikhail A., Kawana M.,
RA   Quertermous E.E., Aoka Y., Fukagawa M., Matsui Y., Platika D.,
RA   Auerbach R., Hogan B.L.M., Snodgrass R., Quertermous T.;
RT   "Cloning and characterization of developmental endothelial locus-1: an
RT   embryonic endothelial cell protein that binds the alphavbeta3 integrin
RT   receptor.";
RL   Genes Dev. 12:21-33(1998).
CC   -!- FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH
CC       INTERACTION WITH THE ALPHA-V BETA3 INTEGRIN RECEPTOR. INHIBITS
CC       FORMATION OF VASCULAR-LIKE STRUCTURES. MAY BE INVOLVED IN
CC       REGULATION OF VASCULAR MORPHOGENESIS OR REMODELING IN EMBRYONIC
CC       DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC   -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS: LONG ISOFORM (SHOWN HERE) AND
CC       SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN ANGIOBLASTS AND EARLY ENDOTHELIAL
CC       CELLS. BY EMBRYONIC DAY 13.5, ALSO EXPRESSED IN A RESTRICTED GROUP
CC       OF NON-ENDOTHELIAL CELLS INCLUDING CHONDROCYTES AND RETINAL
CC       NEURONS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE EMBRYO FROM DAY 7. AFTER DAY
CC       15.5, EXPRESSION DECREASES AND DISAPPEARS COMPLETELY BY THE TIME
CC       OF BIRTH.
CC   -!- SIMILARITY: CONTAINS 3 EGF LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR   EMBL; AF031524; AAB86585.1; -.
DR   EMBL; AF031525; AAB86586.1; -.
DR   HSSP; P07593; 1TFG.
DR   MGD; MGI:1329025; Edil3.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00001; EGF_like; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_3.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   EGF-like domain; Alternative splicing; Signal; Developmental protein;
KW   Cell adhesion; Repeat; Vascularization.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    480       INTEGRIN-BINDING PROTEIN DEL1.
FT   DOMAIN       26     59       EGF-LIKE 1.
FT   DOMAIN       79    116       EGF-LIKE 2.
FT   DOMAIN      123    154       EGF-LIKE 3.
FT   DOMAIN      161    311       F5/8 TYPE C 1.
FT   DOMAIN      322    473       F5/8 TYPE C 2.
FT   SITE         96     98       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     26     37       BY SIMILARITY.
FT   DISULFID     31     48       BY SIMILARITY.
FT   DISULFID     50     59       BY SIMILARITY.
FT   DISULFID     78     89       BY SIMILARITY.
FT   DISULFID     83    105       BY SIMILARITY.
FT   DISULFID    107    116       BY SIMILARITY.
FT   DISULFID    158    314       BY SIMILARITY.
FT   DISULFID    301    305       BY SIMILARITY.
FT   DISULFID    319    476       BY SIMILARITY.
FT   VARSPLIC    218    221       INTQ -> VEVC (IN SHORT ISOFORM).
FT   VARSPLIC    222    480       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   480 AA;  53740 MW;  46D91EFD92617140 CRC64;


  Query Match             83.7%;  Score 36;  DB 11;  Length 480;
  Best Local Similarity   77.8%;  Pred. No. 15;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |||| |||:
Db     402 VQFVGSYKL 410


RESULT   5
Q960N3
ID   Q960N3      PRELIMINARY;      PRT;   330 AA.
AC   Q960N3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LD43270P.
GN   CG11330.
OS   Drosophila melanogaster (Fruit fly)
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y; CN RW SP;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY051966; AAK93390.1; -.
SQ   SEQUENCE   330 AA;  37738 MW;  52F511686617FCD92 CRC64;


  Query Match             74.4%;  Score 32;  DB 5;  Length 330;
  Best Local Similarity   85.7%;  Pred. No. 73;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 QFVASYK 8
           ||||||:
Db     210 QFVASYE 216


RESULT   6
Q9VMA1
ID   Q9VMA1      PRELIMINARY;      PRT;   483 AA.
AC   Q9VMA1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG11330 PROTEIN (CORTEX).
```

```
GN   CG11330 OR CORT.
OS   Drosophila melanogaster (Fruit fly)
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21150807;
RA   Chu T., Henrion G., Haegeli V., Strickland S.;
RT   "Cortex, a Drosophila gene required to complete oocyte meiosis, is a
RT   member of the Cdc20/fizzy protein family.";
RL   Genesis 29:141-152(2001).
CC   -!- SIMILARITY: CONTAINS 2 WD REPEATS (TRP-ASP DOMAINS).
DR   EMBL; AE003614; AAF52421.1; -.
DR   EMBL; AY033478; AAK54464.1; -.
DR   FlyBase; FBgn0000351; cort.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 2.
DR   SMART; SM00320; WD40; 3.
DR   PROSITE; PS50082; WD_REPEATS_2; 2.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Repeat; WD repeat.
SQ   SEQUENCE   483 AA;  55574 MW;  007D00190DE06735 CRC64;

  Query Match             74.4%;  Score 32;  DB 5;  Length 483;
  Best Local Similarity   85.7%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 QFVASYK 8
              ||||||:
Db        363 QFVASYR 369

RESULT 7
Q9VC31
ID   Q9VC31      PRELIMINARY;      PRT;   610 AA.
AC   Q9VC31;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG13638 PROTEIN.
GN   CG13638.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR   EMBL; AE003749; AAF56345.1; -.
DR   HSSP; P05713; 3RAB.
DR   FlyBase; FBgn0039231; CG13638.
DR   InterPro; IPR003579; Rab.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   Pfam; PF00071; ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   SMART; SM00175; RAB; 1.
KW   GTP-binding; Lipoprotein.
SQ   SEQUENCE   610 AA;  64826 MW;  [illegible] CRC64;
```

```
  Query Match             74.4%;  Score 32;  DB 5;  Length 610;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           : |||:|||
Db     400 LDFVATYKV 408


RESULT   8
Q919J2
ID   Q919J2      PRELIMINARY;      PRT;   121 AA.
AC   Q919J2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CUN083 HYPOTHETICAL PROTEIN.
GN   CUN083.
OS   Culex nigripalpus baculovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Baculoviridae.
OX   NCBI_TaxID=130556;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FLORIDA1997;
RX   MEDLINE=21488685; PubMed=11602755;
RA   Afonso C.L., Tulman E.R., Lu Z., Balinsky C.A., Moser B.A.,
RA   Becnel J.J., Rock D.L., Kutish G.F.;
RT   "Genome Sequence of a Baculovirus Pathogenic for Culex nigripalpus.";
RL   J. Virol. 75:11157-11165(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FLORIDA1997;
RA   Afonso C.L., Tulman E.R., Lu Z., Balinsky C.A., Moser B.A.,
RA   Becnel J.J., Rock D.L., Kutish G.F.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF403738; AAK94161.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   121 AA;  13655 MW;  79971FD649E5B78F CRC64;


  Query Match             72.1%;  Score 31;  DB 12;  Length 121;
  Best Local Similarity   66.7%;  Pred. No. 41;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           :|:| |||
Db     105 VEFLALYKV 113


RESULT   9
Q9KTE8
ID   Q9KTE8      PRELIMINARY;      PRT;   213 AA.
AC   Q9KTE8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RARE LIPOPROTEIN B.
GN   VC0954.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-484(2000).
DR   EMBL; AE004177; AAF94116.1; -.
DR   TIGR; VC0954; -.
KW   Lipoprotein, Complete proteome.
SQ   SEQUENCE   213 AA;  24121 MW;  3ECB5E30802C8022 CRC64;


  Query Match             72.1%;  Score 31;  DB 16;  Length 213;
  Best Local Similarity   85.7%;  Pred. No. 75;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 FVASYKV 9
           |||||:|
Db     113 FVASYRV 119


RESULT  10
O94290
ID   O94290      PRELIMINARY;      PRT;   217 AA.
AC   O94290;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE RNA BINDING PROTEIN.
GN   SPBC887.05C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Lyne M., Rajandream M.A., Barrell B.G., Lelaure V., Galibert F.;
RL   Submitted (NOV 1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL033388; CAA21890.1; -.
DR   HSSP; P11940; 1CVJ.
DR   InterPro; IPR000504; RRM.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
SQ   SEQUENCE   217 AA;  25310 MW;  4F42BFA4ABAA0C8F CRC64;


  Query Match             72.1%;  Score 31;  DB 3;  Length 217;
  Best Local Similarity   77.8%;  Pred. No. 76;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 VQFVASYKV 9
           |  ||||||
Db     104 VDHVASYKV 112


RESULT  11
Q9PFS3
ID   Q9PFS3      PRELIMINARY;      PRT;   308 AA.
AC   Q9PFS3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL PROTEIN XF0584.
GN   XF0584.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
```

RA Alvarenga P., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa ";
RL Nature 406:151-159(2000).
DR EMBL; AE003904; AAF83394.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 308 AA; 34464 MW; 9197D8D56D8C8C7ER CRC64;

Query Match 72.1%; Score 31; DB 16; Length 308;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FVASYKV 9
|||||:|
Db 97 FVASYRV 103

RESULT 12
Q9NWI9
ID Q9NWI9 PRELIMINARY; PRT; 310 AA.
AC Q9NWI9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CDNA FLJ20821 FIS, CLONE ADSE00469 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ADIPOSE TISSUE;
RA Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K.,
RA Hirao M., Ohmori Y., Ota T., Suzuki Y., Obayashi M., Nishi T.,
RA Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK000828; BAA91391.1; -.
FT NON_TER 310 310
SQ SEQUENCE 310 AA; 35495 MW; 5CD4332EE6517A9C CRC64;

Query Match 72.1%; Score 31; DB 4; Length 310;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QFVASYK 8
|||||:
Db 19 QFVASYQ 25

RESULT 13
O66257
ID O66257 PRELIMINARY; PRT; 344 AA.
AC O66257;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INTEGRAL MEMBRANE ACETYLTRANSFERASE
OS Actinobacillus actinomycetemcomitans (Haemophilus
OS actinomycetemcomitans).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC9710;
RX MEDLINE 99023763; PubMed=9805802;
RA Nakano Y., Yoshida Y., Yamashita Y., Koga T.;
RT "A gene cluster for 6-deoxy-D-talan synthesis in Actinobacillus
RT actinomycetemcomitans.",
RL Biochim. Biophys. Acta 1442:409-414(1998).
DR EMBL; AR010415; BAA28139.1; -.
DR InterPro; IPR002656; DUF33.
DR InterPro; IPR001680; WD40.
DR Pfam; PF01757; DUF33; 2.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Transferase.
SQ SEQUENCE 344 AA; 40036 MW; DD7B85875EAE8B80 CRC64;

Query Match 72.1%; Score 31; DB 2; Length 344;
Best Local Similarity 62.5%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 QFVASYKV 9
|| : :|||
Db 202 QFIVTYKV 209

RESULT 14
Q91ZJ6
ID Q91ZJ6 PRELIMINARY; PRT; 1206 AA
AC Q91ZJ6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA0903-LIKE PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Fuchs S., Resch K., Thiel C., Platzer M., Rockusch H.,
RA Schmitt-John T.;
RT "An inversion between mouse and man on chromosome 11/2p13-15 detected
RT by high-resolution comparative mapping.";
RL Submitted (SEP 2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF424697; AAL24806.1; -.
SQ SEQUENCE 1206 AA; 136046 MW; 3A9A9A7C7CE0AC45 CRC64;

Query Match 72.1%; Score 31; DB 11; Length 1206;
Best Local Similarity 85.7%; Pred. No. 4.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QFVASYK 8
|||||:
Db 19 QFVASYQ 25

```
RESULT 15
Q58836
ID   Q58836      PRELIMINARY;      PRT;  1226 AA.
AC   Q58836;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ1441.
GN   MJ1441.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: STRONG TO P.DENITRIFICANS COBN AND M.JANNASCHII
CC       MJ0907.
DR   EMBL; U67585; AAB99452.1; -.
DR   TIGR; MJ1441; -.
DR   InterPro; IPR003672; CobN/Mg_chelatase.
DR   Pfam; PF02514; cobN-Mg_chel; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   1226 AA;  141327 MW;  8F7DDF4E36162AB9 CRC64;


  Query Match             72.1%;  Score 31;  DB 17;  Length 1226;
  Best Local Similarity   71.4%;  Pred. No. 5e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 QFVASYK 8
           |::||||
Db     510 QYIASYK 516



Search completed: September  5, 2002, 15:31:07
Job time: 472 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:27:36 ; Search time 58.86 Seconds
(without alignments)
3.735 Million cell updates/sec

Title: US-09-744-804-41
Perfect score: 44
Sequence: 1 EIIAALCGA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/7B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	100.0	387	2	US-08-162-402B-6	Sequence 6, Appli
2	44	100.0	415	2	US-08-162-402B-8	Sequence 8, Appli
3	37	84.1	463	2	US-08-162-402B-9	Sequence 9, Appli
4	36	81.8	355	3	US-08-890-719-11	Sequence 11, Appl
5	36	81.8	355	3	US-08-890-719-13	Sequence 13, Appl
6	32	72.7	361	1	US-08-415 751-36	Sequence 36, Appl
7	31	70.5	15	1	US-08-190-802A-1	Sequence 1, Appli
8	31	70.5	15	4	US-08-477-346-1	Sequence 1, Appli
9	31	70.5	15	4	US-08-473-089-1	Sequence 1, Appli
10	31	70.5	88	4	US-09-216 295-17	Sequence 17, Appl
11	31	70.5	470	2	US-08 691 814B-2	Sequence 2, Appli
12	31	70.5	644	1	US-08-336 708A-9	Sequence 9, Appli
13	31	70.5	1210	2	US-08-484-438-7	Sequence 7, Appli
14	31	70.5	1210	2	US-08-475-035-4	Sequence 4, Appli
15	30	68.2	259	4	US-09-216-295 5	Sequence 5, Appli
16	30	68.2	325	4	US-09-194 905-11	Sequence 11, Appl
17	30	68.2	475	4	US-09-233-989-2	Sequence 2, Appli
18	30	68.2	476	4	US-09-233 989-3	Sequence 3, Appli
19	30	68.2	476	4	US-09-233-989 6	Sequence 6, Appli
20	29	65.9	27	4	US-09-142-469 3	Sequence 3, Appli
21	29	65.9	46	2	US-08-691 814B-17	Sequence 17, Appl
22	29	65.9	51	1	US 08-173-102 8	Sequence 8, Appli
23	29	65.9	51	5	PCT-US94-14617-8	Sequence 8, Appli
24	29	65.9	86	2	US 08-421-144A 9	Sequence 9, Appli
25	29	65.9	104	1	US-08 466-033-83	Sequence 83, Appl
26	29	65.9	104	2	US-08-444 733-83	Sequence 83, Appl
27	29	65.9	104	2	US-08-464 134 83	Sequence 83, Appl
28	29	65.9	104	2	US-08-461-361-83	Sequence 83, Appl
29	29	65.9	104	2	US 08 485 910 83	Sequence 83, Appl
30	29	65.9	104	5	PCT-US95 06266-67	Sequence 67, Appl
31	29	65.9	118	4	US-09-627-476-17	Sequence 17, Appl
32	29	65.9	122	1	US-07-956-700B-37	Sequence 37, Appl
33	29	65.9	122	1	US-08-476-537-37	Sequence 37, Appl
34	29	65.9	122	1	US 08 485 607 37	Sequence 37, Appl
35	29	65.9	122	2	US 08 475-879-37	Sequence 37, Appl
36	29	65.9	125	4	US 09 372-422A 6	Sequence 6, Appli
37	29	65.9	129	2	US-08-785-065-3	Sequence 3, Appli
38	29	65.9	132	4	US 08 676 444 47	Sequence 47, Appl
39	29	65.9	143	2	US-08-611-757-105	Sequence 105, App
40	29	65.9	143	5	PCT-US95-05980-105	Sequence 105, App
41	29	65.9	150	4	US-09-372-422A-46	Sequence 46, Appl
42	29	65.9	159	2	US-08-828-832 1	Sequence 1, Appli
43	29	65.9	190	1	US-08-466-033-38	Sequence 38, Appl
44	29	65.9	190	2	US-08-444-733-38	Sequence 38, Appl
45	29	65.9	190	2	US-08-464-134-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-08-162-402B-6
; Sequence 6, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT. LAROCCA, DAVID J.
; TITLE OF INVENTION. 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION GLYCOLIPID (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES. 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE. Fraser, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER. US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amsel, Viviana
; REGISTRATION NUMBER. 30,930
; REFERENCE/DOCKET NUMBER. P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213 622-7700
; TELEFAX: 213 489 4210
; TELEX:
; INFORMATION FOR SEQ ID NO. 6.
; SEQUENCE CHARACTERISTICS.
; LENGTH: 387 amino acids
; TYPE. amino acid
; STRANDEDNESS. unknown
; TOPOLOGY: unknown
; MOLECULE TYPE. peptide
US-08-162-402B-6

Query Match 100.0%, Score 44, 100.0, Length 387;
Best Local Similarity 100.0%, Prob. No. 0.75,

```
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RLLAALCGA 9
           |||||||||
Db       5 RLLAALCGA 13


RESULT   2
US-08-162-402B-8
; Sequence 8, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  465 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-8


  Query Match             100.0%;  Score 44;  DB 2;  Length 465;
  Best Local Similarity   100.0%;  Pred. No. 0.9;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RLLAALCGA 9
           |||||||||
Db       5 RLLAALCGA 13


RESULT   3
US-08-162-402B-9
; Sequence 9, Application US/08162402B
; Patent No. 5972337
;  GENERAL INFORMATION:
;    APPLICANT:  CERIANI, ROBERTO L.
;    APPLICANT:  PETERSON, JERRY A.
;    APPLICANT:  LAROCCA, DAVID J.
;    TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
;    TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder & Poplawski
;      STREET:  444 South Flower St., 19th Floor
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,402B
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amzel, Viviana
;      REGISTRATION NUMBER:  30,930
;      REFERENCE/DOCKET NUMBER:  P66 38215
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  463 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-162-402B-9


  Query Match             84.1%;  Score 37;  DB 2;  Length 463;
  Best Local Similarity   87.5%;  Pred. No. 17;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RLLAALCG 8
           |:|||||
Db       5 RVLAALCG 12


RESULT   4
US-08-890-719-11
; Sequence 11, Application US/08890719A
; Patent No. 6075125
; GENERAL INFORMATION:
;  APPLICANT: Bacon, Larry D
;  APPLICANT: Hunt, Henry D
;  APPLICANT: Fulton, Janet
;  TITLE OF INVENTION: Production of Antisera Specific to Major
;  TITLE OF INVENTION: Histocompatibility Complex Molecules in Chickens
;  FILE REFERENCE: Dkt 0064.96 - Larry D. Bacon et al.
;  CURRENT APPLICATION NUMBER: US/08/890,719A
;  CURRENT FILING DATE: 1997-07-09
;  EARLIER APPLICATION NUMBER: 60/021,685
;  EARLIER FILING DATE: 1996-07-10
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
;   LENGTH: 355
;   TYPE: PRT
;   ORGANISM: Gallus gallus
US-08-890-719-11


  Query Match             81.8%;  Score 36;  DB 3;  Length 355;
```

```
  Best Local Similarity   87.5%;  Pred. No. 20;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
            ||||:|||
Db      12 LLAAVCGA 19


RESULT   5
US-08-890-719-13
; Sequence 13, Application US/08890719A
; Patent No. 6075125
; GENERAL INFORMATION:
;  APPLICANT: Bacon, Larry D
;  APPLICANT: Hunt, Henry D
;  APPLICANT: Fulton, Janet
;  TITLE OF INVENTION: Production of Antisera Specific to Major
;  TITLE OF INVENTION: Histocompatibility Complex Molecules in Chickens
;  FILE REFERENCE: Dkt 0064.96 - Larry D. Bacon et al.
;  CURRENT APPLICATION NUMBER: US/08/890,719A
;  CURRENT FILING DATE: 1997-07-09
;  EARLIER APPLICATION NUMBER: 60/021,685
;  EARLIER FILING DATE: 1996-07-16
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 355
;   TYPE: PRT
;   ORGANISM: Gallus gallus
US-08-890-719-13


  Query Match             81.8%;  Score 36;  DB 3;  Length 355;
  Best Local Similarity   87.5%;  Pred. No. 20;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
            ||||:|||
Db      12 LLAAVCGA 19


RESULT   6
US-08-415-751-36
; Sequence 36, Application US/08415751
; Patent No. 5644772
;  GENERAL INFORMATION:
;    APPLICANT:  PETERSEN, CAROLYN
;    APPLICANT:  LEECH, JAMES
;    APPLICANT:  NELSON, RICHARD, C.
;    APPLICANT:  GUT, JIRI
;    TITLE OF INVENTION:  POLYPEPTIDES BINDING ANTI-
;    TITLE OF INVENTION:  CRYPTOSPORIDIUM ANTIBODIES, DNA
;    TITLE OF INVENTION:  AND RNA ENCODING THEM, HYBRID
;    TITLE OF INVENTION:  VECTOR AND TRANSFORMED HOST AND
;    TITLE OF INVENTION:  METHODS FOR IMMUNOTHERAPY AND
;    TITLE OF INVENTION:  DIAGNOSIS AND KIT
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PHILLIPS, MOORE, LEMPIO & FINLEY
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.5 inch, 1.44 Kb storage
;      COMPUTER:  PC
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/415,751
;      FILING DATE:  03-APR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/071,880
;      FILING DATE:  June 1, 1993
;      APPLICATION NUMBER:  07/891,301
;      FILING DATE:  May 29, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hana Dolezalova
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.19-2 (HHD)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-1677
;      TELEFAX:  (415) 324-1678
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    ORIGINAL SOURCE:
;      ORGANISM:  Cryptosporidium parvum
;    FEATURE:
;      NAME/KEY:  Positions coded by nonsense codons are
;      NAME/KEY:  identified as Xaa.
US-08-415-751-36


  Query Match             72.7%;  Score 32;  DB 1;  Length 361;
  Best Local Similarity   62.5%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 RLLAALCG 8
            |:|| :||
Db      44 RILAGICG 51


RESULT   7
US-08-190-802A-1
; Sequence 1, Application US/08190802A
; Patent No. 5519003
;  GENERAL INFORMATION:
;    APPLICANT:  Mochly-Rosen, Daria
;    APPLICANT:  Ron, Dorit
;    TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  265
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  P.O. Box 60850
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-0850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/190,802A
;      FILING DATE:  01-FEB-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fabian, Gary R.
;      REGISTRATION NUMBER:  33,875
;      REFERENCE/DOCKET NUMBER:  8600.0139
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15 amino acids
;      TYPE:  amino acid
```

```
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       INDIVIDUAL ISOLATE:  Peptide I
US-08-190-802A-1


  Query Match             70.5%;  Score 31;  DB 1;  Length 15;
  Best Local Similarity   62.5%;  Pred. No. 7.5;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 RLLAALCG 8
            ::| ||||
Db       6 KILVALCG 13


RESULT   8
US-08-477-346-1
; Sequence 1, Application US/08477346
; Patent No. 6262023
;  GENERAL INFORMATION:
;     APPLICANT:  Mochly-Rosen, Daria
;     APPLICANT:  Ron, Dorit
;     TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
;     TITLE OF INVENTION:  Thereof
;     NUMBER OF SEQUENCES:  265
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Morrison & Foerster
;       STREET:  2000 Pennsylvania Avenue, NW
;       CITY:  Washington
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20006-1812
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/477,346
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/487,072
;       FILING DATE:  07-JUN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MURASHIGE, KATE H.
;       REGISTRATION NUMBER:  29,959
;       REFERENCE/DOCKET NUMBER:  2550-0025 20
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 887-1500
;       TELEFAX:  (202) 887-0763
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  15 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       INDIVIDUAL ISOLATE:  Peptide I
US-08-477-346-1


  Query Match             70.5%;  Score 31;  DB 4;  Length 15;
  Best Local Similarity   62.5%;  Pred. No. 7.5;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 RLLAALCG 8
            ::| ||||
Db       6 KILVALCG 13


RESULT   9
US-08-473-089-1
; Sequence 1, Application US/08473089
; Patent No. 6342368
;  GENERAL INFORMATION:
;     APPLICANT:  Mochly-Rosen, Daria
;     APPLICANT:  Ron, Dorit
;     TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
;     TITLE OF INVENTION:  Thereof
;     NUMBER OF SEQUENCES:  265
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Morrison & Foerster
;       STREET:  2000 Pennsylvania Avenue, NW
;       CITY:  Washington
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20006-1812
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/473,089
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MURASHIGE, KATE H.
;       REGISTRATION NUMBER:  29,959
;       REFERENCE/DOCKET NUMBER:  2550-0025 22
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 887-1500
;       TELEFAX:  (202) 887-0763
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  15 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       INDIVIDUAL ISOLATE:  Peptide I
US-08-473-089-1


  Query Match             70.5%;  Score 31;  DB 4;  Length 15;
  Best Local Similarity   62.5%;  Pred. No. 7.5;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 RLLAALCG 8
            ::| ||||
Db       6 KILVALCG 13


RESULT  10
US-09-216-295-17
; Sequence 17, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
;  APPLICANT:  Mitchinson, Colin
;  APPLICANT:  Wendt, Dan J.
;  TITLE OF INVENTION:  No. 6268328el Variant EGIII-Like Cellulase Compositions
;  FILE REFERENCE:  GC555
;  CURRENT APPLICATION NUMBER:  US/09/216,295
;  CURRENT FILING DATE:  1998-12-18
;  NUMBER OF SEQ ID NOS:  41
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 17
;   LENGTH:  88
```

```
;   TYPE: PRT
;   ORGANISM: Gliocladium roseum (3)
US-09-216-245-17


  Query Match             70.5%;  Score 31;  DB 4;  Length 88;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 LAALCG 8
           ||||||
Db      16 LAALCG 21


RESULT  11
US-08-691-814B-2
; Sequence 2, Application US/08691814B
; Patent No. 5981218
;  GENERAL INFORMATION:
;    APPLICANT:  Rio, Marie-Christine
;    APPLICANT:  Tomasetto, Catherine
;    APPLICANT:  Basset, Paul
;    APPLICANT:  Byrne, Jennifer
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Useful
;    TITLE OF INVENTION:  as Leukemia Markers and in Breast Cancer Prognosis
;    NUMBER OF SEQUENCES:  124
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Ave, NW, Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,814B
;      FILING DATE:  31-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/002,183
;      FILING DATE:  09-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Steffe, Eric K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1383.0090001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  470 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-691-814B-2


  Query Match             70.5%;  Score 31;  DB 2;  Length 470;
  Best Local Similarity   75.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 RLLAALCG 8
           |||  |||
Db      15 RLLCPLCG 22


RESULT  12
US-08-336-708A-9
; Sequence 9, Application US/08336708A
; Patent No. 5521295
;  GENERAL INFORMATION:
;    APPLICANT:  Pacifici, Robert E.
;    APPLICANT:  Thomason, Arlen R.
;    APPLICANT:  Chang, Ming-Shi
;    TITLE OF INVENTION:  Hybrid Receptor Molecules
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amgen Inc.
;      STREET:  1840 Dehavilland Drive
;      CITY:  Thousand Oaks
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91320-1789
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/336,708A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oleski, Nancy
;      REFERENCE/DOCKET NUMBER:  A-241A
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  644 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-336-708A-9


  Query Match             70.5%;  Score 31;  DB 1;  Length 644;
  Best Local Similarity   87.5%;  Pred. No. 3.1e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
           ||||||  |
Db      14 LLAALCPA 21


RESULT  13
US-08-484-438-7
; Sequence 7, Application US/08484438
; Patent No. 5811098
; Patent No. 5811098 5780031
;  GENERAL INFORMATION:
;    APPLICANT:  Plowman, Gregory D.
;    APPLICANT:  Culouscou, Jean-Michel
;    APPLICANT:  Shoyab, Mohammed
;    APPLICANT:  Siegall, Clay B.
;    APPLICANT:  Hellstr m, Ingegerd
;    APPLICANT:  Hellstr m, Karl E.
;    TITLE OF INVENTION:  HER4 HUMAN RECEPTOR TYROSINE KINASE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,438
```

```
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/323,442
;       FILING DATE:  14-OCT-1994
;       APPLICATION NUMBER:  US 08/150,704
;       FILING DATE:  10-NOV-1993
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/981,165
;       FILING DATE:  24-NOV-1992
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Misrock, S. Leslie
;       REGISTRATION NUMBER:  18,872
;       REFERENCE/DOCKET NUMBER:  5624-230
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-8864/9741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1210 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  protein
US-08-484-438-7


  Query Match             70.5%;  Score 31;  DB 2;  Length 1210;
  Best Local Similarity   87.5%;  Pred. No. 5.7e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
           |||||| |
Db      14 LLAALCPA 21


RESULT  14
US-08-475-035-4
; Sequence 4, Application US/08475035
; Patent No. 5985553
;  GENERAL INFORMATION:
;     APPLICANT:  KING, C. R.
;     APPLICANT:  KRAUS, MATTHIAS H.
;     APPLICANT:  AARONSON, STUART A.
;     TITLE OF INVENTION:  HUMAN GENE RELATED TO BUT DISTINCT FROM
;     TITLE OF INVENTION:  EGF RECEPTOR GENE
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;       STREET:  Suite 1200, 127 Peachtree Street
;       CITY:  Atlanta
;       STATE:  Georgia
;       COUNTRY:  USA
;       ZIP:  30303
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/475,035
;       FILING DATE:  7 Jun 1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Perryman, David G.
;       REGISTRATION NUMBER:  33,438
;       REFERENCE/DOCKET NUMBER:  1414.656
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  404/688-0770
;       TELEFAX:  404/688-9880
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1210 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-475-035-4


  Query Match             70.5%;  Score 31;  DB 2;  Length 1210;
  Best Local Similarity   87.5%;  Pred. No. 5.7e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
           |||||| |
Db      14 LLAALCPA 21


RESULT  15
US-09-216-295-5
; Sequence 5, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
;  APPLICANT:  Mitchinson, Colin
;  APPLICANT:  Wendt, Dan J.
;  TITLE OF INVENTION:  No. 6268328e1 Variant EGIII-Like Cellulase Compositions
;  FILE REFERENCE:  GC555
;  CURRENT APPLICATION NUMBER:  US/09/216,295
;  CURRENT FILING DATE:  1998-12-18
;  NUMBER OF SEQ ID NOS:  41
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 259
;   TYPE: PRT
;   ORGANISM: Aspergillus aculeatus
US-09-216-295-5


  Query Match             68.2%;  Score 30;  DB 4;  Length 259;
  Best Local Similarity   87.5%;  Pred. No. 1.9e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
           ||||| ||
Db       6 LLAALAGA 13


Search completed:  September  5, 2002, 15:27:37
Job time:  322 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:26:32 ; Search time 158.43 Seconds
(without alignments)
6.310 Million cell updates/sec

Title: US-09-744-804-41
Perfect score: 44
Sequence: 1 RLIAALCGA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073496 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold_geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	100.0	9	21	AAY82846	Lactadherin (BA-46
2	44	100.0	387	16	AAP77252	HMFG 46 kDa antige
3	44	100.0	387	21	AAY94453	Human lactadherin
4	37	84.1	426	21	AAY94454	Mouse lactadherin
5	36	81.8	345	9	AAP83149	Probe P10-encoded
6	36	81.8	355	21	AAY94506	Chicken BFIV21 cla
7	36	81.8	355	21	AAY94507	Chicken BFIV12 cla
8	35	79.5	365	22	ABG15914	Novel human diagno
9	34	77.3	160	22	AAU40306	Propionibacterium
10	34	77.3	197	22	AAU41965	Propionibacterium
11	34	77.3	466	22	ABB70578	Drosophila melanog
12	34	77.3	1754	22	ABG19818	Novel human diagno
13	34	77.3	1759	22	AAG09273	Human polypeptide
14	33	75.0	80	22	AAB85537	Human secreted pro
15	33	75.0	106	22	ABG20209	Novel human diagno
16	33	75.0	218	20	AAY04969	Mycobacterium spec
17	33	75.0	218	22	AAB52411	Mycobacterium tube
18	33	75.0	219	20	AAY04970	Mycobacterium spec
19	33	75.0	268	22	AAG64579	Human cathepsin 29
20	33	75.0	415	22	AAB85557	Human secreted pro
21	33	75.0	418	15	AAR60155	Comamonas testoste
22	33	75.0	548	22	ABB63196	Drosophila melanog
23	33	75.0	811	22	ABG20210	Novel human diagno
24	33	75.0	889	22	ABG20216	Novel human diagno
25	33	75.0	927	22	AAG67259	Amino acid sequenc
26	33	75.0	927	22	AAB94880	Human protein sequ
27	32	72.7	49	22	ABB44074	Peptide #11580 enc
28	32	72.7	49	22	ABG26065	Protein #8065 enco
29	32	72.7	49	22	AAM65095	Human brain expres
30	32	72.7	49	22	AAM77799	Human bone marrow
31	32	72.7	49	22	AAM21787	Peptide #8141 enco
32	32	72.7	49	22	AAM28020	Peptide #12057 enc
33	32	72.7	82	22	AAU40067	Propionibacterium
34	32	72.7	113	22	AAU41814	Propionibacterium
35	32	72.7	201	21	AAY86223	Human secreted pro
36	32	72.7	268	20	AAY34703	C. pneumoniae prot
37	32	72.7	316	22	AAU41023	Propionibacterium
38	32	72.7	388	22	AAC81145	Mycobacterium tube
39	32	72.7	458	22	AAU36419	Pseudomonas aeruqi
40	32	72.7	458	22	AAU00967	Pseudomonas aerugi
41	32	72.7	532	22	AAB72393	Rat L-type amino a
42	32	72.7	535	21	AAB43136	Human OFFX OFF2900
43	32	72.7	535	21	AAY71061	Human membrane tra
44	32	72.7	535	22	AAB72395	Human L type amino
45	32	72.7	564	22	ABB11788	Human amino acid t

ALIGNMENTS

```
RESULT 1
AAY82846
ID   AAY82846 standard; peptide; 9 AA.
XX
AC   AAY82846;
XX
DT   19-JUN-2000 (first entry)
XX
DE   Lactadherin (BA 46) peptide fragment (tumour associated antigen).
XX
KW   Tumour associated antigen peptide; TAA; cancer; carcinoma;
KW   treatment, prevention; cure, anti-tumour vaccine; metastases;
KW   breast, bladder, prostate, pancreas, ovary, thyroid, colon,
KW   stomach; carcinoma; MHC Class I; HLA-A2; human;
KW   Major Histocompatibility Complex; uroplakin;
KW   prostate specific antigen, prostate specific membrane antigen;
KW   prostate acid phosphatase, mucin, lactadherin;
KW   teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRIPTO-1.
XX
OS   Homo sapiens.
XX
PN   WO200005723-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-IL00417.
XX
PR   30-JUL-1998;   98IL-0125608.
XX
PA   (YEDA ) YEDA RES & DEV CO LTD.
PA   (BIOT-) BIO-TECHNOLOGY GEN CORP.
XX
PI   Eisenbach L, Carmon L, Tirosh B, Bar-Haim E, Paz A, Fridkin M,
PI   Fitzer-attas C;
```

XX
DR WPI; 2000-205463/18.
XX
PT Tumor associated antigen peptides, especially derived from uroplakin, useful as vaccines to prevent or cure cancers including breast, bladder, prostate, pancreas, ovary, thyroid, colon and stomach -
XX
PS Claim 17; Page 100; 113pp; English.
XX
CC Tumour associated antigen peptides (TAA) may be used for the treatment, prevention and cure of cancer or cancer metastases. The cancer may be breast, bladder, prostate, pancreas, ovary, thyroid, colon, stomach, head or neck cancer or a carcinoma. The tumour associated antigens are presentable to the immune system by HLA-A2 molecules and are generally between 8 to 10 amino acids in length. The amino acids located at positions 2 and 9 of the tumour associated antigens are the anchor residues which participate in the binding to MHC class I molecules, more specifically HLA A2. More tumour associated antigens are described in GENESEQ records AAY82806-Y82882. Those tumour associated antigens described in records AAY82806-Y82824 and AAY82855 YY82869 are derived from Uroplakin, such as Uroplakin II, Uroplakin Ia, Uroplakin III and Uroplakin Ib. Those described in records AAY82825-Y82829 are derived from prostate specific antigen (PSA). Those described in records AAY82830-Y82835 are derived from prostate specific membrane antigen (PSMA). Those described in records YY82836-AAY82839 are derived from prostate acid phosphotase (PAP). Those described in records AAY82840-Y82846 are derived from Lactadherin (BA-46). Those described in records AAY82847 Y82854 are derived from Mucin and those described in records AAY82871-Y82882 are derived from Teratocarcinoma derived growth factor (CRIPTO-1).
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 44; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLLAALCGA 9
||||||||| |
Db 1 rllaalcga 9

RESULT 2
AAR77252
ID AAR77252 standard; Protein; 387 AA.
XX
AC AAR77252;
XX
DT 21-NOV-1995 (first entry)
XX
DE HMFG 46 kDa antigen.
XX
KW HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide; epithelium; tumor; breast cancer; monoclonal antibody; MAb
XX
OS Homo sapiens.
XX
PN WO9515171-A.
XX
PD 08-JUN-1995.
XX
PF 05-DEC-1994; 94WO-US13967.
XX
PR 03-DEC-1993; 93US-0162402.
XX
PA (CANC-) CANCER RES FUND CONTRA COSTA.
XX
PI Ceriani RL, Larocca DJ, Peterson JA;
XX
DR WPI; 1995-215151/28.
DR N-PSDB; AAQ91198.
XX
PT 46 kD apparent molecular weight human milk fat globule antigen -
PT used in assays to determine the presence of a cancerous tumour of
PT epithelial origin, and in a vaccine against neoplastic tumours
XX
PS Claim 6; Page 46 47; 68pp; English.
XX
CC A complete cDNA sequence for the 46 kDa HMFG antigen, a major
CC component of the apical surface of the normal breast epithelial
CC cell, was obtained by PCR and RACE methods. cDNA clones can be
CC used to prepare MAbs for use in immunotherapy, immunohistopathology,
CC prognosis, diagnosis, imaging and therapy. Recombinant antigen can
CC be expressed in prokaryotic or (glycosylated) in eukaryotic cells.
XX
SQ Sequence 387 AA;

Query Match 100.0%; Score 44; DB 16; Length 387;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLLAALCGA 9
||||||||| |
Db 5 rllaalcga 13

RESULT 3
AAY94453
ID AAY94453 standard; Protein; 387 AA.
XX
AC AAY94453;
XX
DT 11-SEP-2000 (first entry)
XX
DE Human lactadherin protein.
XX
KW Human; lactadherin; MGF-E8; anti-tumour; immune response; exosome; dendritic cell.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..23
FT /label= Secretion_signal
FT Protein 24..387
FT /label= Lactadherin
FT Binding-site 46..48
FT /label= Integrin_binding_site
XX
PN EP1004664-A1.
XX
PD 31-MAY-2000.
XX
PF 24-NOV-1998; 98EP-0402925.
XX
PR 24-NOV-1998; 98EP-0402925.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PA (CURI-) INST CURIE.
XX
DR WPI; 2000-352597/31.
DR N-PSDB; AAA27140.
XX
PT Chimeric isolated (human) lactadherin polypeptide that functions as an
PT adaptor of cross-priming to eliminate pathogenic antigens, e.g. in
PT tumors -
XX
PS Example 3; Page 12; 20pp; English.
XX
CC Lactadherin protein was found in exosomes produced by dendritic cells.
CC The protein is involved in the phagocytosis of particulate antigens by
CC dendritic cells. Exosomes produced by dendritic cells exposed to
CC tumour antigens induce potent immune responses. Lactadherin or variants
CC of it may be used in the mediation of an immune response. Variants of

```
CC   lactadherin may be used for inhibition and/or stimulation of the
CC   cross-priming of antigens and stimulation of the phagocytosis of
CC   antigens by dendritic cells. Compositions derived from lactadherin can
CC   also be used to monitor an immune response, more specifically a CTL
CC   (cytotoxic T lymphocyte) response and also to produce CTLs specific
CC   for a selected antigen. The present sequence is the human
CC   lactadherin protein.
XX
SQ   Sequence   387 AA;


  Query Match             100.0%;  Score 44;  DB 21;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 6.7;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 RLLAALCGA 9
          |||||||||
Db      5 rllaalcga 13


RESULT   4
AAY94454
ID   AAY94454 standard; Protein; 426 AA.
XX
AC   AAY94454;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Mouse lactadherin protein.
XX
KW   Human; lactadherin; MGF E8, anti-tumour; immune response;
KW   exosome; dendritic cell.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1  22
FT                   /label= Secretion_signal
FT   Protein         23..426
FT                   /label= Lactadherin
FT   Binding-site    87  89
FT                   /label= Integrin_binding_site
FT   Misc-difference 93  111
FT                   /note= "encoded by GT"
XX
PN   EP1004664-A1
XX
PD   31-MAY-2000.
XX
PF   24-NOV-1998;   98EP-0402925.
XX
PR   24-NOV-1998;   98EP-0402925.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE
PA   (CURI-) INST CURIE.
XX
DR   WPI; 2000-352597/31.
DR   N-PSDB; AAA27141.
XX
PT   Chimeric isolated (human) lactadherin polypeptide that functions as an
PT   adaptor of cross-priming to eliminate pathogenic antigens, e.g. in
PT   tumors -
XX
PS   Disclosure; Page 12; 20pp. English.
XX
CC   Lactadherin protein was found in exosomes produced by dendritic cells.
CC   The protein is involved in the phagocytosis of particulate antigens by
CC   dendritic cells. Exosomes produced by dendritic cells exposed to
CC   tumour antigens induce potent immune responses. Lactadherin or variants
CC   of it may be used in the mediation of an immune response. Variants of
CC   lactadherin may be used for inhibition and/or stimulation of the
CC   cross priming of antigens and stimulation of the phagocytosis of
CC   antigens by dendritic cells. Compositions derived from lactadherin can
CC   also be used to monitor an immune response, more specifically a CTL
CC   (cytotoxic T-lymphocyte) response and also to produce CTLs specific
CC   for a selected antigen. The present sequence is the mouse
CC   lactadherin protein.
XX
SQ   Sequence   426 AA;


  Query Match             84.1%;  Score 37;  DB 21;  Length 426;
  Best Local Similarity   87.5%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 RLLAALCG 8
          |:||||||
Db      5 rvlaalcg 12


RESULT   5
AAP83149
ID   AAP83149 standard; protein; 345 AA.
XX
AC   AAP83149;
XX
DT   27-NOV-1990  (first entry)
XX
DE   Probe F10 encoded protein of MHC class I of chicken.
XX
KW   chicken erythrocyte; histocompatibility complex class I antigen;
KW   restriction fragment length polymorphism (RFLP); probe F10.
XX
OS   Gallus gallus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label=signal peptide
FT                   /note="cleaved from mature protein"
FT   Domain          23..110
FT                   /label=alpha 1
FT   Domain          111..201
FT                   /label=alpha 2
FT   Domain          202..292
FT                   /label=alpha 3
FT   Region          293..334
FT                   /label=transmembrane
FT   Region          335..345
FT                   /label=cytoplasmic extremity
XX
PN   WO8809386-A.
XX
PD   01-DEC-1988.
XX
PF   27 MAY-1988;   88WO FR00270.
XX
PR   29-MAY-1987;   87FR-0007577.
XX
PA   (DVAL/) DE LA VALLEE BS.
XX
PI   Auffray C,  Behar G,  Billault A,  Chausse AM,  Guillemot F,  Bourlet Y;
XX
DR   WPI; 1988-353959/49.
DR   N-PSDB; AAN80603.
XX
PT   Restriction fragment length polymorphism genotyping test -
PT   for chickens, using erythrocyte DNA fragments and hybridisation
PT   probe derived from histocompatibility complex antigen.
XX
PS   Disclosure; ; p; French.
XX
CC   Probe F10 is used in RFLP analysis of chicken erythrocyte derived
CC   DNA. It corresponds to a MHC class I antigen and was obtained from
CC   the total mRNA of different tissues of chicken.
CC   See also AAN80602 and AAN82429.
XX
```

```
SQ   Sequence   345 AA;

  Query Match             81.8%;  Score 36;  DB 9;  Length 345;
  Best Local Similarity   87.5%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
            ||||:|||
Db      12 llaavcga 19


RESULT   6
AAY94506
ID   AAY94506 standard; Protein; 355 AA.
XX
AC   AAY94506;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Chicken BFIV21 class I MHC protein.
XX
KW   Chicken; MHC; major histocompatibility complex; BFIV;
KW   antisera.
XX
OS   Gallus gallus
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= signal_peptide
FT   Domain          22..109
FT                   /label= alpha1_domain
FT   Domain          110..200
FT                   /label= alpha2_domain
FT   Domain          201..291
FT                   /label= alpha3_domain
FT   Domain          292..355
FT                   /label= transmembrane_and_cytoplasmic
FT   Modified-site   58
FT                   /label= N-glycosylation_site
FT   Modified-site   106
FT                   /label= N-glycosylation_site
XX
PN   US6075125-A
XX
PD   13-JUN-2000
XX
PF   09-JUL-1997;   97US-0890719
XX
PR   10-JUL-1996;   96US-0021685.
XX
PA   (USDA ) US SEC OF AGRIC.
XX
PI   Hunt HD,  Bacon LD,  Fulton JE;
XX
DR   WPI; 2000-411285/35.
DR   N-PSDB; AAA48669.
XX
PT   Producing antisera specific to major histocompatibility complex (MHC)
PT   proteins in chickens involves administering transfected cells
PT   expressing heterologous chicken MHC class I protein capable of
PT   eliciting immune response -
XX
PS   Disclosure; Fig 4; 40pp; English
XX
CC   The chicken Major Histocompatibility Complex (MHC) B-complex is
CC   comprised of three classes of loci. Class I was mutated by
CC   site directed mutagenesis. Transfected cells containing the mutant
CC   sequence may be generated  The heterologous BFIV protein produced by
CC   these cells may be used as an immunogen to produce chicken MHC class I
CC   specific antisera. This antisera may then be used to determine the
CC   BF haplotype of any chicken. BFIV specific antisera may be used to
CC   determine the B haplotype of chickens with reduced cross reaction
CC   with class I and class IV MHC proteins. The present sequence is
CC   BFIV21 class I MHC protein.
XX
SQ   Sequence   355 AA;

  Query Match             81.8%;  Score 36;  DB 21;  Length 355;
  Best Local Similarity   87.5%;  Pred. No. 1.5e+02;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
            ||||:|||
Db      12 llaavcga 19


RESULT   7
AAY94507
ID   AAY94507 standard; Protein; 355 AA.
XX
AC   AAY94507;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Chicken BFIV12 class I MHC protein.
XX
KW   Chicken; MHC; major histocompatibility complex; BFIV;
KW   antisera.
XX
OS   Gallus gallus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= signal_peptide
FT   Misc-difference 3
FT                   /note  "encoded by TCG"
FT   Domain          22..109
FT                   /label= alpha1_domain
FT   Modified-site   58
FT                   /label= N-glycosylation_site
FT   Modified-site   106
FT                   /label= N-glycosylation_site
FT   Domain          110..200
FT                   /label= alpha2_domain
FT   Domain          201..291
FT                   /label= alpha3_domain
FT   Domain          292..355
FT                   /label= transmembrane_and_cytoplasmic
XX
PN   US6075125-A.
XX
PD   13-JUN-2000.
XX
PF   09-JUL-1997;   97US-0890719.
XX
PR   10-JUL-1996;   96US-0021685.
XX
PA   (USDA ) US SEC OF AGRIC.
XX
PI   Hunt HD,  Bacon LD,  Fulton JE;
XX
DR   WPI; 2000-411285/35.
DR   N-PSDB; AAA48670.
XX
PT   Producing antisera specific to major histocompatibility complex (MHC)
PT   proteins in chickens involves administering transfected cells
PT   expressing heterologous chicken MHC class I protein capable of
PT   eliciting immune response -
XX
PS   Disclosure; Fig 4; 40pp; English.
XX
CC   The chicken Major Histocompatibility Complex (MHC) B complex is
CC   comprised of three classes of loci. Class I was mutated by
CC   site directed mutagenesis. Transfected cells containing the mutant
```

CC sequence may be generated. The heterologous BFIV protein produced by
CC these cells may be used as an immunogen to produce chicken MHC class I
CC specific antisera. This antisera may then be used to determine the
CC BF haplotype of any chicken. BFIV specific antisera may be used to
CC determine the B haplotype of chickens with reduced cross reaction
CC with class I and class IV MHC proteins. The present sequence is
CC BFIV12 class I MHC protein.
XX
SQ Sequence 355 AA;

Query Match 81.8%; Score 36; DB 21; Length 355;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLAALCGA 9
||||:|||
Db 12 llaavcga 19

RESULT 8
ABG15914
ID ABG15914 standard; Protein; 365 AA.
XX
AC ABG15914;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #15905.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217
PR 23-AUG-2000; 2000US-0644167.
XX
PA (HYSE-) HYSEQ INC
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-630362/72
DR N-PSDB; AAS80101.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 46273; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity,
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG30810-ABG30977 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 365 AA;

Query Match 79.5%; Score 35; DB 22; Length 365;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLAALCG 8
|||||||
Db 85 llaalcg 91

RESULT 9
AAU40306
ID AAU40306 standard; Protein; 160 AA.
XX
AC AAU40306;
XX
DT 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #1202.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PF 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skolk; YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59511.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris
XX
PS Example 1; SEQ ID No 1501; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to

CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 160 AA;

Query Match 77.3%; Score 34; DB 22; Length 160;
Best Local Similarity 66.7%; Pred No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RLLAALCGA 9
|| :|:|||
Db 17 rlasaicga 25

RESULT 10
AAU41965
ID AAU41965 standard; Protein; 197 AA.
XX
AC AAU41965;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #2861
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-198047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59515.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1, SEQ ID No 3168; 136pp; English
XX
CC Sequences AAU34135-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 197 AA;

Query Match 77.3%; Score 34; DB 22; Length 197;
Best Local Similarity 75.0%; Pred No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLAALCGA 9
::|||||
Db 52 ivaalcga 59

RESULT 11
ABB70578
ID ABB70578 standard; Protein; 466 AA.
XX
AC ABB70578;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 38526.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL14681.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 38526; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01846-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 466 AA;

```
  Query Match             77.3%;  Score 34;  DB 22;  Length 466;
  Best Local Similarity   75.0%;  Pred. No. 4.2e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 RLLAALCG 8
           |||| :||
Db     122 rllaqicg 129


RESULT  12
ABG18818
ID   ABG18818 standard; Protein; 1754 AA.
XX
AC   ABG18818;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #18809.
XX
KW   Human, chromosome mapping, gene mapping, gene therapy, forensic,
KW   food supplement; medical imaging; diagnostic; genetic disorder
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PP   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0644167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS83005.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID NO 49177; 193pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (I) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence   1754 AA;

  Query Match             77.3%;  Score 34;  DB 22;  Length 1754;
  Best Local Similarity   77.8%;  Pred. No. 1.5e+03;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 RLLAALCGA 9
           |||| || |
Db    1587 rllatlcsa 1595


RESULT  13
AA009273
ID   AA009273 standard; Protein; 1759 AA.
XX
AC   AA009273;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 23165.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI89204.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 23165; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AA179941-AA193841) and
CC   the encoded proteins (AA009010-AA010910) that exhibit activity relating to
CC   cytokines, cell proliferation or cell differentiation and which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1759 AA;

  Query Match             77.3%;  Score 34;  DB 22;  Length 1759;
  Best Local Similarity   77.8%;  Pred. No. 1.5e+03;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 RLLAALCGA 9
           |||| || |
Db    1590 rllatlcsa 1598
```

```
RESULT  14
AAB85537
ID   AAB85537 standard; protein; 80 AA.
XX
AC   AAB85537;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human secreted protein (clone id HCHPU32).
XX
KW   secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW   antiproliferative, cytostatic, cardiant, vasotropic, cerebroprotective;
KW   nootropic, neuroprotective, antibacterial, virucide, fungicide, human,
KW   opthalmalogical; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 11
FT                   /note= "Xaa can be any amino acid"
XX
PN   WO200155430-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01431.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628
PR   12-SEP-2000; 2000US-0231968.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Komatsoulis GA,  Baker KP,  Birse CE,  Soppet DR,  Olsen HS,
PI   Moore PA,  Wei P,  Ebner R,  Duan DR,  Shi Y,  Choi GH,  Fiscella M;
PI   Ni J,  Ruben SM,  Barash SC;
XX
DR   WPI; 2001-476228/51.
DR   N-PSDB; AAH46947.
XX
PT   17 isolated nucleic acid molecules encoding human secreted proteins,
PT   used to preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 11; Page 454; 482pp; English.
XX
CC   The invention provides novel human secreted proteins and polynucleotides
CC   encoding them. The secreted proteins can be expressed by standard
CC   recombinant methodology. The secreted proteins and polynucleotides are
CC   used to prevent, treat or ameliorate a medical condition in e.g. humans,
CC   mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They can
CC   also be used in diagnosing a pathological condition. The antibodies to
CC   the proteins can also be used in alleviating symptoms associated with the
CC   disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme
CC   linked immunosorbent assays (ELISA). Disorders which are diagnosed or
CC   treated include autoimmune diseases e.g. rheumatoid arthritis,
CC   hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC   cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC   e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
CC   Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC   ocular disorders e.g. corneal infection. The polypeptides can also be
CC   used to aid wound healing and epithelial cell proliferation, to prevent
CC   skin aging due to sunburn, to maintain organs before transplantation, for
CC   supporting cell culture of primary tissues, to regenerate tissues and in
CC   chemotaxis. The polypeptides can also be used as a food additive or
CC   preservative to increase or decrease storage capabilities. The present
CC   sequence represents a human secreted protein.
XX
SQ   Sequence   80 AA;

  Query Match             75.0%;  Score 33;  DB 22;  Length 80;
  Best Local Similarity   85.7%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 LAALCGA 9
           :||||||
Db      43 maalcga 49


RESULT  15
ABG20209
ID   ABG20209 standard; Protein; 106 AA.
XX
AC   ABG20209;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #20200.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217
PR   23-AUG-2000; 2000US-0644167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73
DR   N-PSDB; AAS84396
XX
PI   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20, SEQ ID No 50568, 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00018-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   106 AA;

  Query Match             75.0%;  Score 33;  DB 22;  Length 106;
  Best Local Similarity   85.7%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;
```

Qy 3 LAALCGA 9
:||||||
Db 82 maalcga 88

Search completed: September 5, 2002, 15:26:33
Job time: 373 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:28:54 ; Search time 69.78 Seconds
(without alignments)
12.393 Million cell updates/sec

Title: US-09-744-804 41
Perfect score: 41
Sequence: 1 RLLAALCCA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	37	84.1	427	2	JC4915	aqs protein precur
2	37	84.1	463	1	A36479	milk fat globule m
3	36	81.8	141	2	S62512	hypothetical prote
4	36	81.8	287	2	B83299	hypothetical prote
5	36	81.8	345	1	HLCHB4	MHC class I histoc
6	36	81.8	355	2	I51409	major histocompati
7	36	81.8	355	2	T28149	MHC class I histoc
8	36	81.8	355	2	T28152	MHC class I histoc
9	35	79.5	230	2	D87446	hypothetical prote
10	34	77.3	171	2	C87418	hypothetical prote
11	34	77.3	376	2	F75503	conserved hypothet
12	34	77.3	457	2	A10955	probable PTS syste
13	34	77.3	1744	1	C4HU	complement C4A pre
14	33	75.0	196	2	AD0385	4-methyl-5(B-hydro
15	33	75.0	218	2	F70524	hypothetical prote
16	33	75.0	235	2	T04860	hypothetical prote
17	33	75.0	307	2	T19582	hypothetical prote
18	33	75.0	334	2	T04520	probable nicotinat
19	33	75.0	462	1	S30104	indoleacetamide hy
20	33	75.0	466	2	I39707	probable indoleace
21	33	75.0	467	1	HJA3I	indoleacetamide hy
22	33	75.0	467	1	Q2ACAT	indoleacetamide hy
23	33	75.0	467	2	AH3228	indole acetamide h
24	33	75.0	469	2	T25526	probable integral
25	33	75.0	644	2	A36325	epidermal growth f
26	33	75.0	3161	2	T30442	protein HMWP1 - Ye
27	33	75.0	4717	2	T41581	hypothetical colle
28	32	72.7	102	2	T20283	hypothetical prote
29	32	72.7	159	1	I80299	sigma E factor req
30	32	72.7	159	2	D91058	sigma-E factor req
31	32	72.7	159	2	B85903	sigma-E factor, ne
32	32	72.7	205	2	G72531	hypothetical prote
33	32	72.7	244	2	C97532	probable transcrip
34	32	72.7	244	2	AC2751	transcription regu
35	32	72.7	274	2	G71426	hypothetical prote
36	32	72.7	305	2	T04558	ribose transport A
37	32	72.7	313	2	H72777	probable cobD - My
38	32	72.7	325	2	AC1047	lysine--tRNA ligas
39	32	72.7	325	2	G92207	hypothetical prote
40	32	72.7	325	2	G92224	probable vitamin B
41	32	72.7	344	2	H86502	cytochrome oxidase
42	32	72.7	344	2	G72119	cytochrome D ubiqu
43	32	72.7	355	2	I51152	major histocompati
44	32	72.7	356	2	E75266	CCHFF family prote
45	32	72.7	357	2	I45291	probable nicotinat

ALIGNMENTS

RESULT 1
JC4915
aqs protein precursor - rat
N;Alternate names: O-acetyl-GD3 ganglioside
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
C;Accession: JC4915
R;Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A;Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside
A;Reference number: JC4915; MUID:96374422
A;Accession: JC4915
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-427 <OGU>
A;Cross-references: DDBJ:D84068; NID:g1620986; PIDN:BAA12210.1; PID:g1620987
A;Experimental source: CST cell
C;Comment: This protein is required for the O acetylation of disialoganglioside siali
C;Genetics:
A;Gene: aqs
C;Superfamily: milk fat globule protein; discoidin I amino terminal homology; EGF hom
F;1-21/Domain: signal sequence #status predicted <SIG>
F;28-60/Domain: EGF homology <EG1>
F;68-107/Domain: EGF homology <EG2>
F;119-267/Domain: discoidin I amino terminal homology <DN1>
F;271-427/Domain: discoidin I amino terminal homology <DN2>

Query Match 84.1%; Score 37; DB 2; Length 427;
Best Local Similarity 87.5%; Pred. No. .29;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 RLLAALCCA 8
            |:|||||||
Db       5 RVLAALCCG 12
```

RESULT 2
A36479
milk fat globule membrane protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A36479
R;Stubbs, J.D.; Lekutis, C.; Singer, K.H.; Bui, A.; Yuzuki, D.; Srinivasan, U.; Parry
Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
A;Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the
A;Reference number: A36479; MUID:91046008
A;Accession: A36479
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-463 <STU>
A;Cross-references: GB:M38037; NID:g199142; PIDN:AAA39534.1; PID:g199143

C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
C;Keywords: membrane protein
F;28-66/Domain: EGF homology <EG1>
F;68-107/Domain: EGF homology <EG2>
F;147-303/Domain: discoidin I amino-terminal homology <DN1>
F;307-463/Domain: discoidin I amino-terminal homology <DN2>

Query Match 84.1%; Score 37; DB 1; Length 463;
Best Local Similarity 87.5%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLLAALCG 8
:|||||||
Db 5 RVLAALCG 12

RESULT 3
D83148
hypothetical protein PA3986 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: D83148
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, E.T.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337
A;Accession: D83148
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-143 <STO>
A;Cross-references: GB:AE004816; GB:AE004091; NID:g9950168; PIDN:AAG07474.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3986

Query Match 81.8%; Score 36; DB 2; Length 143;
Best Local Similarity 77.8%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RLLAALCGA 9
:|| |||||
Db 108 QLLCALCGA 116

RESULT 4
E83398
hypothetical protein PA1977 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: E83398
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, E.T.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337
A;Accession: E83398
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-287 <STO>
A;Cross-references: GB:AE004624; GB:AE004091; NID:g9947073; PIDN:AAG05365.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1977

Query Match 81.8%; Score 36; DB 2; Length 287;
Best Local Similarity 87.5%; Pred. No. 32;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLAALCGA 9
:||||||||
Db 242 VLAALCGA 249

RESULT 5
HLCHB4
MHC class I histocompatibility antigen B-F IV alpha chain precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 30-Jun-1992 #sequence_revision 30-Jan-1992 #text_change 22-Jun-1999
C;Accession: A45846; S01172
R;Kroemer, G.; Zoorob, R.; Auffray, C.
Immunogenetics 31, 405-409, 1990
A;Title: Structure and expression of a chicken MHC class I gene.
A;Reference number: A45846; MUID:90316612
A;Accession: A45846
A;Molecule type: DNA
A;Residues: 1-345 <KRO>
A;Cross-references: GB:M31012
R;Guillemot, F.; Billault, A.; Pourquie, O.; Behar, G.; Chausse, A.M.; Zoorob, R.; Kr
EMBO J. 7, 2775-2785, 1988
A;Title: A molecular map of the chicken major histocompatibility complex: the class I
A;Reference number: S01172; MUID:89030642
A;Accession: S01172
A;Molecule type: mRNA
A;Residues: 1-345 <GUI>
A;Cross-references: EMBL:X12780; NID:g63089; PIDN:CAA31272.1; PID:g63090
C;Genetics:
A;Map position: 16
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology
C;Keywords: glycoprotein; heterodimer; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-345/Product: class I histocompatibility antigen B-F IV alpha chain #status predi
F;23-301/Domain: extracellular #status predicted <EXT>
F;23-110/Domain: alpha-1 <EXT1>
F;111-201/Domain: alpha-2 <EXT2>
F;214-279/Domain: immunoglobulin homology <IMM>
F;302-324/Domain: transmembrane #status predicted <TMM>
F;325-345/Domain: intracellular #status predicted <INT>
F;59,107/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;121-183,221-277/Disulfide bonds: #status predicted

Query Match 81.8%; Score 36; DB 1; Length 345;
Best Local Similarity 87.5%; Pred. No. 37;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLAALCGA 9
||||:|||
Db 12 LLAAVCGA 19

RESULT 6
I51309
major histocompatibility complex class I glycoprotein haplotype B21 - chicken
C;Species: Gallus gallus (chicken)
C;Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 21-Jan-2000
C;Accession: I51309
R;Fulton, J.E.; Thacker, E.L.; Bacon, L.D.; Hunt, H.D.
Eur. J. Immunol. 25, 2069-2076, 1995
A;Title: Functional analysis of avian class I (BF/B) glycoproteins by epitope tagging
A;Reference number: I51309; MUID:95347411
A;Accession: I51309
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-355 <FUL>
A;Cross-references: GB:S78682; NID:g1042200; PIDN:AAB34945.1; PID:g1042201
C;Genetics:
A;Gene: BF/B21
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology
F;213-278/Domain: immunoglobulin homology <IMM>

Query Match 81.8%; Score 36; DB 2; Length 355;
Best Local Similarity 87.5%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLAALCGA 9
||||:|||
Db 12 LLAAVCGA 19

RESULT 7
T28149
MHC class I histocompatibility antigen B-F alpha chain 2 - chicken
C;Species: Gallus gallus (chicken)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T28149
R;Milne, S.; Kaufman, J.; Beck, S.
submitted to the EMBL Data Library, May 1998
A;Description: DNA sequencing and analysis of the chicken major histocompatibility compl
A;Reference number: Z20475
A;Accession: T28149
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-355 <MIL>
A;Cross-references: EMBL:AL023516, PIDN:CAA18969.1
A;Experimental source: clone cB12
C;Genetics:
A;Gene: BFa2
A;Map position: 16
A;Introns: 22/1; 110/1; 201/1; 252/1; 328/1; 355/1
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 81.8%; Score 36; DB 2; Length 355;
Best Local Similarity 87.5%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLAALCGA 9
||||:|||
Db 12 LLAAVCGA 19

RESULT 8
T28152
MHC class I histocompatibility antigen B-F IV alpha chain precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T28152
R;Milne, S.; Kaufman, J.; Beck, S.
submitted to the EMBL Data Library, May 1998
A;Description: DNA sequencing and analysis of the chicken major histocompatibility compl
A;Reference number: Z20475
A;Accession: T28152
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-355 <MIL>
A;Cross-references: EMBL:AL023516; PIDN:CAA18972.1
A;Experimental source: clone cB12
C;Genetics:
A;Gene: BFa1
A;Map position: 16
A;Introns: 22/1; 110/1; 201/1; 252/1; 328/1; 355/1
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 81.8%; Score 36; DB 2; Length 355;
Best Local Similarity 87.5%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLAALCGA 9
||||:|||
Db 12 LLAAVCGA 19

RESULT 9
D87446
hypothetical protein CC1589 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: D87446
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kon, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: D87446
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-230 <STO>
A;Cross-references: GB:AE005673; NID:g13422092; PIDN:AAK23089.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1589

Query Match 79.5%; Score 35; DB 2; Length 230;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLAALCGA 9
|||| |||
Db 6 LLAAFCGA 13

RESULT 10
C87418
hypothetical protein CC1362 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: C87418
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kon, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: C87418
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-171 <STO>
A;Cross-references: GB:AE005673; NID:g13422713; PIDN:AAK23443.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1362

Query Match 77.3%; Score 34; DB 2; Length 171;
Best Local Similarity 87.5%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLAALCGA 9
|||| |||
Db 19 LLAAACGA 26

RESULT 11
F75503
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75503
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896

A;Accession: F75503
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-376 <WHI>
A;Cross-references: GB:AE001914; GB:AE000513; NID:g6458253; PIDN:AAF10140.1; PID:g6458253
A;Experimental source: strain P1
C;Genetics:
A;Gene: DR0560
A;Map position: 1

Query Match 77.3%; Score 34; DB 2; Length 376;
Best Local Similarity 88.9%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLLAALCGA 9
|||||| ||
Db 38 RLLAALLGA 46

RESULT 12
A10955
probable PTS system permease [imported] - Salmonella enterica subsp. enterica serovar Ty
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C;Accession: A10955
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Maule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AR0502; MUID:21677608
A;Accession: A10955
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-457 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD03141.1; PID:g16504776; GSPDB:GN00176
C;Genetics:
A;Gene: STY3925

Query Match 77.3%; Score 34; DB 2; Length 457;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALCGA 9
:|||:|||
Db 204 MLASLCGA 211

RESULT 13
C4HU
complement C4A precursor [validated] - human
N;Contains: classical complement pathway C3/C5 convertase (EC 3.4.21.43); C4b subunit; C4
C;Species: Homo sapiens (man)
C;Date: 25-Feb-1985 #sequence_revision 23-Aug-1996 #text_change 08-Dec-2000
C;Accession: I56095; A29177; B29177; A90845; A19311; A92337; S12866; A17265; A32335; A27
R;Yu, C.Y.
J. Immunol. 146, 1057-1066, 1991
A;Title: The complete exon-intron structure of a human complement component C4A gene. DN
A;Reference number: I56095; MUID:91108039
A;Accession: I56095
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1744 <PFS>
A;Cross-references: GB:M59815; NID:g179672; PIDN:AAA51855.1; PID:g179674
R;Belt, K.T.; Yu, C.Y.; Carroll, M.C.; Porter, R.R.
Immunogenetics 21, 173-180, 1985
A;Title: Polymorphism of human complement component C4
A;Reference number: A29177; MUID:85156269
A;Accession: A29177
A;Molecule type: DNA
A;Residues: 1-22 <BEL>
A;Cross-references: GB:M14823
A;Accession: B29177
A;Molecule type: DNA
A;Residues: 1056-1225 <BE3>
A;Cross-references: GB:M14824; NID:g179675; PIDN:AAA52292.1; PID:g553210
R;Belt, K.T.; Carroll, M.C.; Porter, R.R.
Cell 36, 907-914, 1984
A;Title: The structural basis of the multiple forms of human complement component C4.
A;Reference number: A90845; MUID:84156544
A;Accession: A90845
A;Molecule type: mRNA
A;Residues: 29-446,'S',348-417,'A',419-725,'P',727-1200,'S',1202-1285,'S',1287-1418,1
A;Cross-references: GB:K02403
R;Carroll, M.C.; Porter, R.R.
Proc. Natl. Acad. Sci. U.S.A. 80, 264-267, 1983
A;Title: Cloning of a human complement component C4 gene.
A;Reference number: A19311; MUID:83117835
A;Accession: A19311
A;Molecule type: mRNA
A;Residues: 1195-1285,'S',1287-1294 <CAR>
A;Cross-references: GB:V00503; GB:J00080; NID:g30010; PIDN:CAA23760.1; PID:g1335040
R;Moon, K.E.; Gorski, J.P.; Hugli, T.E.
J. Biol. Chem. 256, 8685-8692, 1981
A;Title: Complete primary structure of human C4a anaphylatoxin.
A;Reference number: A92337; MUID:81264286
A;Accession: A92337
A;Molecule type: protein
A;Residues: 680-725,'TN',728-756 <MOO>
R;Hessing, M.; van't Veer, C.; Hackeng, T.M.; Bouma, B.N.; Iwanaga, S.
FEBS Lett. 271, 131-136, 1990
A;Title: Importance of the alpha(3)-fragment of complement C4 for the binding with C4
A;Reference number: S12866; MUID:91032049
A;Accession: S12866
A;Molecule type: protein
A;Residues: 757,'X',759-771;980-990 <HES>
R;Campbell, R.D.; Gagnon, J.; Porter, R.R.
Biochem. J. 199, 359-370, 1981
A;Title: Amino acid sequence around the thiol and reactive acyl groups of human compl
A;Reference number: A17265; MUID:82182029
A;Accession: A17265
A;Molecule type: protein
A;Residues: 957-1012,'E',1014-1044 <CAM>
R;Chakravarti, D.N.; Campbell, R.D.; Porter, R.R.
Mol. Immunol. 24, 1187-1197, 1987
A;Title: The chemical structure of the C4d fragment of the human complement component
A;Reference number: A32335; MUID:88094444
A;Accession: A32335
A;Molecule type: protein
A;Residues: 957-1012,'E',1014-1108,'I',1110-1175,'S',1177-1270,'V',1272-1336 <CHA>
A;Note: 1073-Gly, 1120-Leu, 1121-Ser, 1124-Ile, 1125-His, 1207-Ala, 1210-Arg were als
R;Chakravarti, D.N.; Campbell, R.D.; Gagnon, J.
FEBS Lett. 154, 387-390, 1983
A;Title: Amino acid sequence of a polymorphic segment from fragment C4d of human comp
A;Reference number: A27600; MUID:83158189
A;Accession: A27600
A;Molecule type: protein
A;Residues: 1199-1270,'V',1272-1299,'V',1301-1304 <CH2>
R;Whitehead, A.S.; Goldberger, G.; Woods, D.E.; Markham, A.F.; Colten, H.R.
Proc. Natl. Acad. Sci. U.S.A. 80, 5387-5391, 1983
A;Title: Use of a cDNA clone for the fourth component of human complement (C4) for an
A;Reference number: I58991; MUID:84299979
A;Accession: I58991
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1448-1474 <RE2>
A;Cross-references: GB:X00838; NID:g187772; PIDN:AAA36229.1; PID:g187773
R;Sargent, C.A.; Anderson, M.J.; Hsieh, S.L.; Kendall, F.; Gomez-Escobar, N.; Campbel
Hum. Mol. Genet. 3, 481-488, 1994
A;Title: Characterisation of the novel gene G11 lying adjacent to the complement C4A
A;Reference number: I37396; MUID:94282044
A;Accession: I37399

```
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-21 <RE3>
A;Cross-references: EMBL:X77491; NID:g453410; PIDN:CAA54627.1; PID:g453411
C;Comment: This protein is synthesized as a single-chain precursor and, prior to secreti
by disulfide bonds
C;Comment: The activation of complement C4 by complement subcomponent C1s releases the C
 of complement factor 2 to form the classical-complement-pathway C3 convertase. The C4b,
ay C5 convertase.
C;Comment: C4a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.
C;Comment: The activity of C4b is regulated by proteolytic cleavage involving C4b-bindin
C;Comment: Residues 1446 or 1449 may be the carboxyl end of the alpha chain.
C;Comment: There are at least two genes coding for C1, C4A and C4B. Each gene has many J
C;Genetics:
A;Gene: GDB:C4A
A;Cross-references: GDB:119732; OMIM:120810
A;Map position: 6p21.3-6p21.3
A;Introns: 22/2; 88/3; 156/1; 179/3; [illegible]; 447/3; [illegible]
3; 1226/1; 1303/3; 1359/3; 1379/3; 1411/1; 1473/2; [illegible]
C;Superfamily: alpha-2-macroglobulin
C;Keywords: acute phase; complement classical pathway; glycoprotein; hydrolase; inflamma
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-675/Product: complement C4 beta chain #status predicted <BET>
F;20-675,757-1446,1451-1744/Product: C4b #status predicted <C4B>
F;680-1446/Product: complement C4 alpha chain #status predicted <ALP>
F;680-756/Product: C4a anaphylatoxin #status experimental <C4A>
F;757-845/Region: C4b-binding protein binding
F;957-1336/Product: C4d fragment #status experimental <C4D>
F;1454-1744/Product: complement C4 gamma chain #status predicted <GAM>
F;756-757/Cleavage site: Arg-Ala (complement subcomponent C1s) #status experimental
F;1010-1013/Cross-link: thiolester (Cys-Gln) #status experimental
F;1328/Binding site: carbohydrate (Asn) (covalent) #status experimental


  Query Match             77.3%;  Score 34;  DB 1;  Length 1744;
  Best Local Similarity   77.8%;  Pred. No. 3.3e+02;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 RLLAALCGA 9
            |||| || |
Db     1577 RLLATLCSA 1585

RESULT 14
AD0385
4-methyl-5(B-hydroxyethyl)-thiazol monophosphate biosynthesis enzyme [imported] - Yersin
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C;Accession: AD0385
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AD0385
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-196 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC92407.1; PID:g15981110; GSPDB:GN00175
C;Genetics:
A;Gene: thiJ
C;Superfamily: signal transduction protein DJ-1


  Query Match             75.0%;  Score 33;  DB 2;  Length 196;
  Best Local Similarity   66.7%;  Pred. No. 82;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 RLLAALCGA 9
            ||:||:| |
Db      100 RLVAAICAA 108
```

```
RESULT 15
F70524
hypothetical protein Rv0309 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 28-Jul-2000
C;Accession: F70524
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
, Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: F70524
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-218 <[illegible]>
A;Cross-references: [illegible]
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0309
C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0309


  Query Match             75.0%;  Score 33;  DB 2;  Length 218;
  Best Local Similarity   77.8%;  Pred. No. 90;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 RLLAALCGA 9
            |||| || |
Db        3 RLLALLCAA 11


Search completed: September 5, 2002, 15:28:55
Job time: 360 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:43 ; Search time 33.99 Seconds
(without alignments)
10.252 Million cell updates/sec

Title: US-09-744-804-41
Perfect score: 44
Sequence: 1 RLLAALCGA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	100.0	387	1	MFGM_HUMAN	Q08431 homo sapien
2	37	84.1	427	1	MFGM_RAT	P79490 rattus norv
3	37	84.1	463	1	MFGM_MOUSE	P21956 mus musculu
4	36	81.8	287	1	GLPS_PSEAE	P29370 pseudomonas
5	36	81.8	345	1	HA1F_CHICK	P15979 gallus gall
6	34	77.3	879	1	MANB_HUMAN	O00462 homo sapien
7	34	77.3	1744	1	CO4_HUMAN	P01028 homo sapien
8	33	75.0	245	1	CBIM_SALTY	Q05594 salmonella
9	33	75.0	267	1	THIM_PASMU	P57931 pasteurella
10	33	75.0	334	1	COBI_RHOCA	Q52679 rhodobacter
11	33	75.0	462	1	HYI1_AGRVI	Q04557 agrobacteri
12	33	75.0	466	1	HYIN_AGRRH	Q09102 agrobacteri
13	33	75.0	467	1	HYI2_AGRVI	P25016 agrobacteri
14	33	75.0	467	1	HYIN_AGRT4	P03868 agrobacteri
15	32	72.7	159	1	RSEC_ECOLI	P46187 escherichia
16	32	72.7	313	1	COBD_MYCTU	Q10518 mycobacteri
17	32	72.7	325	1	SYK3_SALTY	Q9zj12 salmonella
18	32	72.7	357	1	CORT_STRCO	Q93211 streptomyce
19	32	72.7	363	1	MRS1_SACDO	P41905 saccharomyc
20	32	72.7	389	1	THIL_MYCTU	O06629 mycobacteri
21	32	72.7	393	1	THIL_MYCLE	P46707 mycobacteri
22	32	72.7	422	1	U183_HUMAN	Q6bsa1 homo sapien
23	32	72.7	422	1	U183_MOUSE	Q922r1 mus musculu
24	32	72.7	422	1	U183_RAT	Q08654 rattus norv
25	32	72.7	444	1	TRFD_ARATH	Q02166 arabidopsis
26	32	72.7	455	1	SYN_MYCPN	P75521 mycoplasma
27	32	72.7	456	1	SYN_MYCGE	P47359 mycoplasma
28	32	72.7	531	1	LAT2_MOUSE	Q91x69 mus musculu
29	32	72.7	533	1	LAT2_RAT	Q9wvr6 rattus norv
30	32	72.7	535	1	LAT2_HUMAN	Q9uhi5 homo sapien
31	32	72.7	1876	1	PPSA_MYCTU	Q10977 mycobacteri
32	32	72.7	2280	1	CCAC_SCHPO	P78820 schizosacch
33	31	70.5	196	1	THIJ_ECOLI	Q46948 escherichia
34	31	70.5	196	1	THIJ_SALTY	P55880 salmonella
35	31	70.5	251	1	YAI8_SCHPO	Q10178 schizosacch
36	31	70.5	299	1	TIPW_PEA	P25794 pisum sativ
37	31	70.5	296	1	METF_AQUAE	O67422 aquifex aeo
38	31	70.5	305	1	YH24_ARCFU	O28550 archaeoglob
39	31	70.5	330	1	PEPG_ECOLI	P24177 escherichia
40	31	70.5	330	1	YDEZ_ECOLI	P77651 escherichia
41	31	70.5	345	1	ANX1_HUMAN	P04083 homo sapien
42	31	70.5	345	1	ANX1_MOUSE	P10107 mus musculu
43	31	70.5	345	1	ANX1_RAT	P07150 rattus norv
44	31	70.5	346	1	ANX1_BOVIN	P46193 bos taurus
45	31	70.5	346	1	ANX1_CAVCU	P14087 cavia cutle

ALIGNMENTS

RESULT 1
MFGM_HUMAN
ID MFGM_HUMAN STANDARD; PRT; 387 AA.
AC Q08431;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMFG)
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
GN MFGE8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast, and Breast carcinoma;
RX MEDLINE=96213908; PubMed=8639264;
RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
RT "Cloning and sequence analysis of human breast epithelial antigen
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal
RT growth factor-like domain.";
RL DNA Cell Biol. 15:281-286(1996).
RN [2]
RP SEQUENCE OF 170-387 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=91371451; PubMed=1909932;
RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
RA Ceriani R.L.;
RT "A Mr 46,000 human milk fat globule protein that is highly expressed
RT in human breast tumors contains factor VIII-like domains.";
RL Cancer Res. 51:4994-4998(1991).
RN [3]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=98194921; PubMed=9535276;
RA Giuffrida M.G., Cavaletto M., Giunta C., Conti A.,
RA Godovac-Zimmermann J.;
RT "Isolation and characterization of full and truncated forms of human
RT breast carcinoma protein BA46 from human milk fat globule membranes.";
RL J. Protein Chem. 17:143-148(1998).
RN [4]
RP SEQUENCE OF 254-387, AND IDENTIFICATION OF MEDIN.
RX MEDLINE=99342076; PubMed=10411933;
RA Hanayama R., Nagata S., Slattery R., Westermark G.T., Marchiano G.,
RA Tjernberg L.O., Nordstedt C., Engstroem U., Westermark P.;
RT "Medin: an internal fragment of aortic smooth muscle cell-produced
RT lactadherin forms the most common human amyloid.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=97407085; PubMed=9260929;
RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
RT (RGD) dependent cell adhesion.";

```
RL   DNA Cell Biol 16:861-869(1997)
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
CC       SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
CC   -!- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
CC       MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC   -!- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
CC       AMINO ACID 264 AND 273.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U58516; AAC50549.1; -.
DR   EMBL; S56151; AAB19771.1; -.
DR   MIM; 602281; -.
DR   InterPro; IPR000561; EGF-like
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid
FT   SIGNAL       1     23
FT   CHAIN       24    387       LACTADHERIN.
FT   CHAIN      282    387       LACTADHERIN, SHORT FORM
FT   CHAIN      268    317       MEDIN.
FT   DOMAIN      24     67       EGF-LIKE.
FT   DOMAIN      70    225       F5/8 TYPE C 1.
FT   DOMAIN     230    387       F5/8 TYPE C 2.
FT   SITE        46     48       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    27     38       BY SIMILARITY.
FT   DISULFID    32     55       BY SIMILARITY.
FT   DISULFID    57     66       BY SIMILARITY.
FT   DISULFID    70    225       BY SIMILARITY.
FT   DISULFID   212    216       BY SIMILARITY.
FT   DISULFID   230    387       BY SIMILARITY.
FT   CARBOHYD   238    238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   325    325       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   329    329       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   350    350       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   387 AA;  43123 MW;  2EEB571DEC83782D CRC64;


  Query Match             100.0%;  Score 44;  DB 1;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 0.57;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RILAALCGA 9
           |||||||||
Db       5 RILAALCGA 13


RESULT   2
MFGM_RAT
ID   MFGM_RAT       STANDARD;      PRT;   427 AA.
AC   P70490;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (O-
DE   acetyl GD3 ganglioside synthase) (AGS) (MFGM).
GN   MFGE8 OR AGS
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata, Craniata, Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia, Sciurognathi; Muridae, Murinae, Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98274422; PubMed=8780713;
RA   Ogura K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
RT   "Cloning and expression of cDNA for O-acetylation of GD3
RT   ganglioside.";
RL   Biochem. Biophys. Res. Commun. 225:932-938(1996).
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO
CC       PARTICIPATE IN THE O-ACETYLATION OF GD3 GANGLIOSIDE SIALIC ACID.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D84068; BAA12210.1; -.
DR   HSSP; P00740; 1IXA.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001438; EGF_1I.
DR   InterPro; IPR000421; FA58_C
DR   Pfam; PF00008; EGF; 2
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   PRINTS; PR00010; EGFBLOOD
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Repeat; EGF-like domain; Milk
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23    427       LACTADHERIN.
FT   DOMAIN      24     61       EGF-LIKE 1.
FT   DOMAIN      64    108       EGF-LIKE 2
FT   DOMAIN     111    267       F5/8 TYPE C 1.
FT   DOMAIN     272    427       F5/8 TYPE C 2.
FT   DISULFID    28     39       BY SIMILARITY
FT   DISULFID    33     49       BY SIMILARITY.
FT   DISULFID    51     60       BY SIMILARITY.
FT   DISULFID    68     79       BY SIMILARITY.
FT   DISULFID    73     96       BY SIMILARITY.
FT   DISULFID    98    107       BY SIMILARITY.
FT   DISULFID   111    267       BY SIMILARITY.
FT   DISULFID   254    258       BY SIMILARITY.
FT   DISULFID   272    427       BY SIMILARITY.
FT   SITE        87     89       CELL ATTACHMENT SITE (POTENTIAL).
FT   CARBOHYD    61     61       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   230    230       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   280    280       N-LINKED (GLCNAC     ) (POTENTIAL).
FT   CARBOHYD   390    390       N-LINKED (GLCNAC     ) (POTENTIAL).
SQ   SEQUENCE   427 AA;  47413 MW;  EA8C8631F3EE6047 CRC64;


  Query Match             84.1%;  Score 37;  DB 1;  Length 427;
  Best Local Similarity   87.5%;  Pred. No. 11;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RILAALCG 8
           |:||||||
Db       5 RVLAALCG 12
```

```
RESULT   3
MFGM_MOUSE
ID   MFGM_MOUSE     STANDARD;      PRT;   463 AA.
AC   P21956; P97800;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM)
DE   (Sperm surface protein SP47) (MP47).
GN   MFGE8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.
RC   TISSUE=Mammary gland;
RX   MEDLINE=91046008; PubMed=2122462;
RA   Stubbs J.D., Lekutis C., Singer K.L., Bui A., Yuzuki D.,
RA   Srinivasan U., Parry G.;
RT   "cDNA cloning of a mouse mammary epithelial cell surface protein
RT   reveals the existence of epidermal growth factor-like domains linked
RT   to factor VIII-like sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
RN   [2]
RP   SEQUENCE OF 23-463 FROM N.A.
RC   TISSUE=Testis;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC       SPERMATOZOAN.
CC   -!- DEVELOPMENTAL STAGE: MRNA EXPRESSION IS DETECTABLE IN MAMMARY
CC       TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M38337; AAA39534.1; -.
DR   EMBL; Y11684; CAA72380.1; -.
DR   PIR; A36479; A36479.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:102768; Mfge8.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001438; EGF_11.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein, Repeat, EGF like domain, Milk.
FT   SIGNAL        1     22
FT   CHAIN        23    463       LACTADHERIN
FT   DOMAIN       24     61       EGF-LIKE 1.
FT   DOMAIN       64    108       EGF-LIKE 2.
FT   DOMAIN      148    303       F5/8 TYPE C 1.
FT   DOMAIN      308    463       F5/8 TYPE C 2.
FT   SITE         87     89       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     28     39       BY SIMILARITY.
FT   DISULFID     33     49       BY SIMILARITY.
FT   DISULFID     51     60       BY SIMILARITY.
FT   DISULFID     68     79       BY SIMILARITY.
FT   DISULFID     73     96       BY SIMILARITY.
FT   DISULFID     98    107       BY SIMILARITY.
FT   DISULFID    148    303       BY SIMILARITY.
FT   DISULFID    290    294       BY SIMILARITY.
FT   DISULFID    308    463       BY SIMILARITY.
FT   CARBOHYD     61     61       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    266    266       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    316    316       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    426    426       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT     30     30       S -> F (IN REF. 2).
FT   CONFLICT     35     35       N -> D (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    110    147       ETNYYNLDGEYMFTTAVPNTAVPTPAPTPDLSNNLASR ->
FT                                G (IN REF. 2).
FT   CONFLICT    168    168       Y -> S (IN REF. 2).
FT   CONFLICT    196    196       H -> T (IN REF. 2).
FT   CONFLICT    369    369       L -> S (IN REF. 2).
FT   CONFLICT    395    395       E -> A (IN REF. 2).
SQ   SEQUENCE   463 AA;  51405 MW;  D7B86C6EFBBA724D CRC64;


  Query Match             84.1%;  Score 37;  DB 1;  Length 463;
  Best Local Similarity   87.5%;  Pred. No. 12;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RLLAALCG 8
           |:||||||
Db       5 RVLAALCG 12


RESULT   4
GLPS_PSEAE
ID   GLPS_PSEAE     STANDARD;      PRT;   287 AA.
AC   P29370;
DT   01-DEC-1992 (Rel. 24, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sensor protein glpS (EC 2.7.3.-).
GN   GLPS OR PA1977.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria, gamma subdivision, Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437347; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
RN   [2]
RP   SEQUENCE OF 1-77 FROM N.A.
RX   MEDLINE=92041962; PubMed=1778886;
RA   Schweizer H.P.;
RT   "The agmR gene, an environmentally responsive gene, complements
RT   defective glpE, which encodes the putative activator for glycerol
RT   metabolism in Pseudomonas aeruginosa.";
RL   J. Bacteriol. 173:6798-6806(1991).
CC   -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM GLPS/GLPR
CC       INVOLVED IN THE REGULATION OF GLYCEROL METABOLISM. GLPS MAY
CC       FUNCTION AS A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT
CC       PHOSPHORYLATES GLPR IN RESPONSE TO ENVIRONMENTAL SIGNALS.
CC   -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
```

```
CC       KINASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004624; AAG05465.1; -.
DR   EMBL; M60805; AAA25828.1; -.
DR   InterPro; IPR000620; DUF6.
DR   Pfam; PF00892; DUF6; 2.
KW   Sensory transduction; Transferase; Kinase; Phosphorylation;
KW   Complete proteome.
SQ   SEQUENCE   287 AA;  [illegible]
```

Query Match 81.8%; Score 36; DB 1; Length 287;
Best Local Similarity 87.5%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       2 LLAALCGA 9
           :||||||| 
Db     242 VLAALCGA 249
```

```
RESULT   5
HA1F_CHICK
ID   HA1F_CHICK     STANDARD;      PRT;   345 AA.
AC   P15979;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Class I histocompatibility antigen, F10 alpha chain precursor (B-F
DE   histocompatibility F10 antigen) (B-F-beta-IV) (B12).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B12; TISSUE=Liver;
RX   MEDLINE=89030642; PubMed=3141149;
RA   Guillemot F., Billault A., Pourquie O., Behar G., Chausse A.M.,
RA   Zoorob R., Kreibich G., Auffray C.;
RT   "A molecular map of the chicken major histocompatibility complex: the
RT   class II beta genes are closely linked to the class I genes and the
RT   nucleolar organizer.";
RL   EMBO J. 7:2775-2785(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90316612; PubMed=2370987;
RA   Kroemer G., Zoorob R., Auffray C.;
RT   "Structure and expression of a chicken MHC class I gene.";
RL   Immunogenetics 31:405-409(1990).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X12780; CAA31272.1; -.
DR   EMBL; M31012; AAA48947.1; -.
DR   PIR; S01172; HLCHB4.
DR   PIR; A45846; A45846.
DR   HSSP; P03989; 1HSA.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003597; Ig_c1.
DR   InterPro; IPR001039; MHC_I.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00129; MHC_I; 1.
DR   ProDom; PD000050; MHC_I; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL       1     22
FT   CHAIN       23    345       CLASS I HISTOCOMPATIBILITY ANTIGEN, F10
FT                               ALPHA CHAIN.
FT   DOMAIN       23    110       EXTRACELLULAR ALPHA-1.
FT   DOMAIN      111    202       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      202    292       EXTRACELLULAR ALPHA-3.
FT   DOMAIN      293    301       CONNECTING PEPTIDE.
FT   TRANSMEM    302    324
FT   DOMAIN      325    345       CYTOPLASMIC.
FT   DISULFID    121    183       BY SIMILARITY.
FT   DISULFID    221    277       BY SIMILARITY.
FT   CARBOHYD     59     59       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    107    107       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    339    339       P -> PDRECGSSSSST (IN REF. 2).
SQ   SEQUENCE   345 AA;  38246 MW;  430DCCF8091B69A4 CRC64;
```

Query Match 81.8%; Score 36; DB 1; Length 345;
Best Local Similarity 87.5%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       2 LLAALCGA 9
           ||||:||| 
Db      12 LLAAVCGA 19
```

```
RESULT   6
MANB_HUMAN
ID   MANB_HUMAN     STANDARD;      PRT;   879 AA.
AC   O00462; Q9NYX9; Q96BC3;
DT   15-JUL-1998 (Rel. 36, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Beta-mannosidase precursor (EC 3.2.1.25) (Mannanase) (Mannase).
GN   MANBA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98046009; PubMed=9384606;
RA   Alkhayat A.H., Kraemer S.A., Leipprandt J.R., Macek M., Kleijer W.J.,
RA   Friderici K.H.;
RT   "Human beta mannosidase cDNA characterization and first identification
RT   of a mutation associated with human beta-mannosidosis.";
RL   Hum. Mol. Genet. 7:75-83(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Chang H.-M., Tsai S.-F.;
RT   "Genome sequencing of the chromosome 4q region implicated in human
RT   hepatocellular carcinoma pathogenesis.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: EXOGLYCOSIDASE THAT CLEAVES THE SINGLE BETA-LINKED
CC       MANNOSE RESIDUE FROM THE NON-REDUCING END OF ALL N-LINKED
CC       GLYCOPROTEIN OLIGOSACCHARIDES.
```

```
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non reducing beta-D-
CC       mannose residues in beta D-mannosides.
CC   -!- PATHWAY: PENULTIMATE STEP IN N LINKED OLIGOSACCHARIDE CATABOLISM.
CC   -!- SUBCELLULAR LOCATION: Lysosomal.
CC   -!- DISEASE: DEFECTS IN MANBA ARE THE CAUSE OF A MILD DISORDER THAT
CC       AFFECTS PERIPHERAL AND CENTRAL NERVOUS SYSTEM MYELIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U60337; AAC39573.1; -.
DR   EMBL; AF224669; AAF35233.1; -.
DR   EMBL; BC015743; AAH15743.1; -.
DR   MIM; 248510; -
DR   InterPro; IPR001649; Glyco_hydro_2.
DR   Pfam; PF02836; Glyco_hydro_2_C; 1.
KW   Hydrolase; Glycosidase; Glycoprotein; Lysosome; Signal.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    879       BETA-MANNOSIDASE.
FT   ACT_SITE   457    457       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    28     28       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    35     35       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    77     77       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   280    280       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   284    284       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   297    297       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   302    302       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   763    763       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT    70     70       Y -> H (IN REF. 1).
FT   CONFLICT   701    701       T -> M (IN REF. 2).
SQ   SEQUENCE   879 AA;  100860 MW;  526D79770CFAD742 CRC64;


  Query Match             77.3%;  Score 34;  DB 1;  Length 879;
  Best Local Similarity   87.5%;  Pred. No. 74;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
           || ||||
Db       7 LLLALCGA 14


RESULT   7
CO4_HUMAN
ID   CO4_HUMAN      STANDARD;      PRT;  1744 AA.
AC   P01028; Q9UIP5; Q9NPK5; Q13906; Q13160; Q14835;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Complement C4 precursor [Contains: C4A anaphylatoxin].
GN   C4A AND C4B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (C4A AND C4B).
RC   TISSUE=Liver;
RX   MEDLINE=84156544; PubMed=6546707;
RA   Belt K.T., Carroll M.C., Porter R.R.;
RT   "The structural basis of the multiple forms of human complement
RT   component C4.";
RL   Cell 36:907-914(1984).
RN   [2]
RP   SEQUENCE FROM N.A. (C4A)
RX   MEDLINE=91108039; PubMed=1988494;
RA   Yu C.Y.;
RT   "The complete exon-intron structure of a human complement component
RT   C4A gene. DNA sequences, polymorphism, and linkage to the
RT   21-hydroxylase gene.";
RL   J. Immunol. 146:1057-1066(1991).
RN   [3]
RP   SEQUENCE FROM N.A. (C4B).
RC   TISSUE=Blood;
RX   MEDLINE=96163032; PubMed=8575841;
RA   Uhlati S., Townsend D.C., Christiansen F.T., Dawkins R.L.,
RA   Abraham L.J.;
RT   "Complete sequence of the complement C4 gene from the HLA-A1, B8,
RT   C4AQ0, C4B1, DR3 haplotype.";
RL   Immunogenetics 43:250-252(1996).
RN   [4]
RP   SEQUENCE FROM N.A. (C4B).
RA   Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA   Banta A., Swartzell S., Smith T.M., Spies T., Hood L.;
RT   "Sequence determination of 300 kilobases of the human class III MHC
RT   locus.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A. (C4A).
RA   Barlow K.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE OF 1-22 AND 1056-1225 FROM N.A.
RX   MEDLINE=85156269; PubMed=3838531;
RA   Belt K.T., Yu C.Y., Carroll M.C., Porter R.R.;
RT   "Polymorphism of human complement component C4.";
RL   Immunogenetics 21:173-180(1985).
RN   [7]
RP   SEQUENCE OF 680-756.
RX   MEDLINE=81264286; PubMed=6167582;
RA   Moon K.E., Gorski J.P., Hugli T.E.;
RT   "Complete primary structure of human C4a anaphylatoxin.";
RL   J. Biol. Chem. 256:8685-8692(1981).
RN   [8]
RP   SEQUENCE OF 957-1044.
RX   MEDLINE=82182029; PubMed=6978711;
RA   Campbell R.D., Gagnon J., Porter R.R.;
RT   "Amino acid sequence around the thiol and reactive acyl groups of
RT   human complement component C4.";
RL   Biochem. J. 199:359-370(1981).
RN   [9]
RP   SEQUENCE OF 990-1037.
RX   MEDLINE=82150875; PubMed=6950484;
RA   Harrison R.A., Thomas M.L., Tack B.F.;
RT   "Sequence determination of the thiolester site of the fourth
RT   component of human complement.";
RL   Proc. Natl. Acad. Sci. U.S.A. 78:7388-7392(1981).
RN   [10]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE=94282044; PubMed=8012361;
RA   Sargent C.A., Anderson M.J., Bsisch S.L., Kendall E.,
RA   Gomez-Escobar N., Campbell R.D.;
RT   "Characterisation of the novel gene C11 lying adjacent to the
RT   complement C4A gene in the human major histocompatibility complex.";
RL   Hum. Mol. Genet. 3:481-488(1994).
RN   [11]
RP   SEQUENCE OF 1405-1431, AND SULFATION.
RX   MEDLINE=86111851; PubMed=3944109;
RA   Hortin G., Sims H., Strauss A.W.;
RT   "Identification of the site of sulfation of the fourth component of
RT   human complement.";
RL   J. Biol. Chem. 261:1786-1793(1986).
RN   [12]
RP   STRUCTURAL BASIS OF POLYMORPHISM.
RX   MEDLINE=87089272; PubMed=2431902;
RA   Yu C.Y., Belt K.T., Giles C.M., Campbell R.D., Porter R.R.;
RT   "Structural basis of the polymorphism of human complement components
RT   C4A and C4B: gene size, reactivity and antigenicity.";
RL   EMBO J. 5:2873-2881(1986).
RN   [13]
```

```
RP   VARIANT C4A6 ALLOTYPE.
RX   MEDLINE=92242905; PubMed=1573268;
RA   Anderson M.J., Milner C.M., Cotton G.H., Campbell R.D.;
RT   "The coding sequence of the hemolytically inactive C4A6 allotype of
RT   human complement component C4 reveals that a single arginine to
RT   tryptophan substitution at beta-chain residue 458 is the likely cause
RT   of the defect.";
RL   J. Immunol. 148:2795-2802(1992).
CC   -!- FUNCTION: C4 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       CLASSICAL PATHWAY OF THE COMPLEMENT SYSTEM. IT IS PROCESSED BY
CC       ACTIVATED C1 WHICH REMOVE FROM THE ALPHA CHAIN THE C4A
CC       ANAPHYLATOXIN.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C4,
CC       C4A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES.
CC   -!- SUBUNIT: THIS PROTEIN IS SYNTHESIZED AS A SINGLE-CHAIN PRECURSOR
CC       AND, PRIOR TO SECRETION, IS ENZYMATICALLY CLEAVED TO FORM A TRIMER
CC       OF NONIDENTICAL CHAINS (ALPHA, BETA, AND GAMMA).
CC   -!- POLYMORPHISM: HUMAN COMPLEMENT COMPONENT C4 IS POLYMORPHIC WITH AT
CC       LEAST TWO LOCI, C4A AND C4B. 13 ALLELES OF C4A AND 22 ALLELES OF
CC       C4B HAVE BEEN DETECTED. THE ALLELE SHOWN HERE IS C4A4.
CC   -!- POLYMORPHISM: THE C4A ALLELES CARRY THE BLOOD GROUP ROGERS WHILE
CC       THE C4B ALLELES CARRY THE BLOOD GROUP CHIDO.
CC   -!- DISEASE: THE C4A6 ALLOTYPE IS TOTALLY DEFICIENT IN HEMOLYTIC
CC       ACTIVITY.
CC   -!- MISCELLANEOUS: C4A ALLOTYPES REACT MORE RAPIDLY WITH THE AMINO
CC       GROUP OF PEPTIDE ANTIGENS WHILE C4B ALLOTYPES REACT MORE RAPIDLY
CC       WITH THE HYDROXYL GROUP OF CARBOHYDRATE ANTIGENS.
CC   -!- MISCELLANEOUS: C4 IS A MAJOR HISTOCOMPATIBILITY COMPLEX CLASS-III
CC       PROTEIN.
CC   -!- SIMILARITY: TO C3, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02403; AAB59537.1; -.
DR   EMBL; K02404; AAA59651.1; -.
DR   EMBL; M59815; AAA51855.1; -.
DR   EMBL; M59816; AAA51855.1; JOINED
DR   EMBL; U24578; AAA99717.1; -.
DR   EMBL; AF019413; AAB67980.1; -.
DR   EMBL; AL049547; CAB89402.1; -.
DR   EMBL; M14823; AAA35617.1; -.
DR   EMBL; M14824; AAA52292.1; -.
DR   EMBL; X77491; CAA54627.1; -.
DR   PIR; A01262; C4HU.
DR   PIR; A17265; A17265.
DR   PIR; A29177; A29177.
DR   PIR; B20807; B20807.
DR   HSSP; P01031; 1KJS.
DR   SWISS-2DPAGE; P01028; HUMAN.
DR   MIM; 120810; -.
DR   MIM; 120820; -.
DR   MIM; 120790; -.
DR   InterPro; IPR002890; A2M_N.
DR   InterPro; IPR001599; Alpha_2_macroglobln.
DR   InterPro; IPR000020; Anaphylatoxin.
DR   InterPro; IPR001840; Anaphylatoxn
DR   InterPro; IPR001134; Netrin_C
DR   Ptam; PF00207; A2M; 1.
DR   Ptam; PF01835; A2M_N; 1.
DR   Ptam; PF01821; ANATO; 1.
DR   Ptam; PF01759; NTR; 1.
DR   PRINTS; PR00004; ANAPHYLATOXN
DR   ProDom; PD003264; Anaphylatoxin; 1
DR   SMART; SM00104; ANATO; 1.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR   PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
DR   PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
KW   Complement pathway; Plasma; Glycoprotein; Sulfation; Signal;
KW   Inflammatory response; Polymorphism; Disease mutation;
KW   Blood group antigen.
FT   SIGNAL       1     19
FT   CHAIN       20    675       COMPLEMENT C4, BETA CHAIN.
FT   PROPEP      676    679
FT   CHAIN       680   1446       COMPLEMENT C4, ALPHA CHAIN.
FT   PROPEP     1447   1453
FT   CHAIN      1454   1744       COMPLEMENT C4, GAMMA CHAIN.
FT   PEPTIDE     680    756       C4A ANAPHYLATOXIN.
FT   DOMAIN      702    736       ANAPHYLATOXIN-LIKE.
FT   DISULFID    702    728       BY SIMILARITY.
FT   DISULFID    703    735       BY SIMILARITY.
FT   DISULFID    716    736       BY SIMILARITY.
FT   THIOLEST   1010   1013
FT   MOD_RES    1417   1417       SULFATION.
FT   MOD_RES    1420   1420       SULFATION.
FT   MOD_RES    1422   1422       SULFATION.
FT   CARBOHYD    226    226       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    862    862       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1328   1328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1391   1391       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     477    477       R -> W (IN C4A6).
FT                                /FTId=VAR_001987.
FT   VARIANT     726    726       P -> L (IN C4A3).
FT                                /FTId=VAR_001988.
FT   VARIANT    1073   1073       D -> G (IN C4A1, C4B1 AND C4B3).
```

Query Match 77.3%; Score 34; DB 1; Length 1744;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 RLLAALCGA 9
           |||| || |
Db    1577 RLLATLCSA 1585


RESULT   8
CBIM_SALTY
ID   CBIM_SALTY     STANDARD;      PRT;   245 AA.
AC   Q05594;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   CbiM protein.
GN   CBIM OR STM2023.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria, gamma subdivision, Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE 93273696; PubMed=8501034;
RA   Roth J.R., Lawrence J.G., Rubenfield M., Kieffer-Higgins S.,
RA   Church G.M.;
RT   "Characterization of the cobalamin (vitamin B12) biosynthetic genes
RT   of Salmonella typhimurium.";
RL   J. Bacteriol. 175:3303-3316(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE 21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
```

```
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- FUNCTION: NOT KNOWN.
CC   -!- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L12006; AAA27264.1; -.
DR   EMBL; AE008789; AAL20927.1; -.
DR   StyGene; SG10045; cbiM.
DR   InterPro; IPR002751; CbiM.
DR   Pfam; PF01891; CbiM; 1.
DR   ProDom; PD005331; CbiM; 1.
KW   Cobalamin biosynthesis; Complete proteome.
SQ   SEQUENCE   245 AA;  26299 MW;  C1B1CC8031E21F20 CRC64;

  Query Match             75.0%;  Score 33;  DB 1;  Length 245;
  Best Local Similarity   66.7%;  Pred. No. 38;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 RLLAALCGA 9
              ::: |||||
Db         72 KVLLALCGA 80

RESULT 9
THIM_PASMU
ID   THIM_PASMU     STANDARD;      PRT;   267 AA.
AC   P57931;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hydroxyethylthiazole kinase (EC 2.7.1.50) (4-methyl-5-beta-
DE   hydroxyethylthiazole kinase) (Thz kinase) (TH kinase).
GN   THIM OR PM1262.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PM70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC   -!- CATALYTIC ACTIVITY: ATP + 4-methyl-5-(2-hydroxyethyl)-thiazole =
CC       ADP + 4-methyl-5-(2-phosphoethyl)-thiazole.
CC   -!- COFACTOR: MAGNESIUM (BY SIMILARITY).
CC   -!- PATHWAY: THIAMINE BIOSYNTHESIS.
CC   -!- SIMILARITY: BELONGS TO THE THZ KINASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE006165; AAK03346.1; -.
DR   InterPro; IPR000417; Hyethyz_kinase.
DR   Pfam; PF02110; HK; 1.
DR   PRINTS; PR01099; HYETHTZKNASE.
KW   Thiamine biosynthesis; Transferase; Kinase; ATP-binding; Magnesium;
KW   Complete proteome.
FT   METAL        91     91       MAGNESIUM (BY SIMILARITY).
FT   METAL       123    123       MAGNESIUM (BY SIMILARITY).
FT   ACT_SITE    194    194       BASE (BY SIMILARITY).
SQ   SEQUENCE   267 AA;  28215 MW;  C710B90C6BB5E971 CRC64;

  Query Match             75.0%;  Score 33;  DB 1;  Length 267;
  Best Local Similarity   75.0%;  Pred. No. 41;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LLAALCGA 9
              ||:|:|||
Db        195 LLSAVCGA 202

RESULT 10
COBT_RHOCA
ID   COBT_RHOCA     STANDARD;      PRT;   334 AA.
AC   Q52679; O68087;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase
DE   (EC 2.4.2.21) (NN:DBI PRT) (N1-alpha-phosphoribosyltransferase).
GN   COBT OR COBU.
OS   Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1061;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SB1003 / ST LOUIS;
RX   MEDLINE=97404404; PubMed=9256491;
RA   Vlcek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;
RT   "Sequence of a 189-kb segment of the chromosome of Rhodobacter
RT   capsulatus SB1003.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).
RN   [2]
RP   SEQUENCE OF 1-116 FROM N.A.
RC   STRAIN=ATCC 33303 / B10;
RX   MEDLINE=95362677; PubMed=7635831;
RA   Pollich M., Klug G.;
RT   "Identification and sequence analysis of genes involved in late steps
RT   in cobalamin (vitamin B12) synthesis in Rhodobacter capsulatus.";
RL   J. Bacteriol. 177:4481-4487(1995).
CC   -!- FUNCTION: CATALYZES THE SYNTHESIS OF ALPHA-RIBAZOLE-5'-PHOSPHATE
CC       FROM NICOTINATE MONONUCLEOTIDE (NAMN) AND 5,6-
CC       DIMETHYLBENZIMIDAZOLE (DMB).
CC   -!- CATALYTIC ACTIVITY: Beta-nicotinate D-ribonucleotide +
CC       dimethylbenzimidazole = nicotinate + N1-(5-phospho-alpha-D-
CC       ribosyl)-5,6-dimethylbenzimidazole.
CC   -!- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC   -!- SIMILARITY: BELONGS TO THE COBT FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF010496; AAC16173.1; -.
DR   EMBL; Z46611; CAA86577.1; -.
DR   HSSP; Q05603; 1D0V.
DR   InterPro; IPR003200; DBI_PRT.
DR   Pfam; PF02277; DBI_PRT; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Glycosyltransferase.
FT   ACT_SITE    303    303       BASE (BY SIMILARITY).
FT   CONFLICT     49     49       A -> S (IN REF. 2).
FT   CONFLICT     66     66       A -> P (IN REF. 2).
SQ   SEQUENCE   334 AA;  34096 MW;  356A74BDEA623637 CRC64;
```

```
  Query Match             75.0%;  Score 33;  DB 1;  Length 334;
  Best Local Similarity   66.7%;  Pred. No. 49;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0,  Gaps    0;

Qy       1 RLLAALCGA 9
           | :||:|||
Db     227 REIAAMCGA 235


RESULT  11
HYI1_AGRVI
ID   HYI1_AGRVI     STANDARD;      PRT;   462 AA.
AC   Q04557;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Indoleacetamide hydrolase (EC 3.5.1.-) (IAH) (Indole-3-acetamide
DE   hydrolase).
GN   IAAH.
OS   Agrobacterium vitis (Rhizobium vitis).
OG   Plasmid pTiS4.
OC   Bacteria; Proteobacteria, alpha subdivision, Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=373;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN S4.
RX   MEDLINE 93101133, PubMed=1465104,
RA   Canaday J., Gerard J.-C., Crouzet P., Otten L.;
RT   "Organization and functional analysis of three T-DNAs from the
RT   vitopine Ti plasmid pTiS4.";
RL   Mol. Gen. Genet. 235:292-303(1992).
CC   -!- CATALYTIC ACTIVITY: INDOLE-3-ACETAMIDE (IAM) = INDOLE-3-
CC       ACETIC ACID (IAA).
CC   -!- PATHWAY: SECOND STEP IN THE BIOSYNTHESIS OF AUXINS FROM
CC       TRYPTOPHAN.
CC   -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M91609; AAA98148.1; -.
DR   InterPro; IPR000120; Amidase.
DR   Pfam; PF01425; Amidase; 1.
DR   PROSITE; PS00571; AMIDASES; 1.
KW   Hydrolase; Auxin biosynthesis; Crown gall tumor; T-DNA; Plasmid.
SQ   SEQUENCE   462 AA;  49810 MW;  A4B78DAB34BBEAE5 CRC64;


  Query Match             75.0%;  Score 33;  DB 1;  Length 462;
  Best Local Similarity   87.5%;  Pred. No. 65;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0,  Gaps    0;

Qy       1 RLLAALCG 8
           || |||||
Db     175 RLPAALCG 182


RESULT  12
HYIN_AGRRH
ID   HYIN_AGRRH     STANDARD;      PRT;   466 AA.
AC   Q09102;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Indoleacetamide hydrolase (EC 3.5.1.-) (IAH) (Indole-3-acetamide
DE   hydrolase).
GN   AUX2.
OS   Agrobacterium rhizogenes
OG   Plasmid pRiA4.
OC   Bacteria; Proteobacteria, alpha subdivision, Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=359;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A4;
RX   MEDLINE=92033088; PubMed=1932811;
RA   Camilleri C., Jouanin L.;
RT   "The TR-DNA region carrying the auxin synthesis genes of the
RT   Agrobacterium rhizogenes agropine type plasmid pRiA4. Nucleotide
RT   sequence analysis and introduction into tobacco plants.";
RL   Mol. Plant Microbe Interact. 4:155-162(1991).
CC   -!- CATALYTIC ACTIVITY: INDOLE-3-ACETAMIDE (IAM) = INDOLE-3-
CC       ACETIC ACID (IAA).
CC   -!- PATHWAY: SECOND STEP IN THE BIOSYNTHESIS OF AUXINS FROM
CC       TRYPTOPHAN.
CC   -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61151; AAA22079.1; -.
DR   InterPro; IPR000120; Amidase.
DR   Pfam; PF01425; Amidase; 1.
DR   PROSITE; PS00571; AMIDASES; 1.
KW   Hydrolase; Auxin biosynthesis; T-DNA; Plasmid.
SQ   SEQUENCE   466 AA;  49533 MW;  7375A5994B903995 CRC64;


  Query Match             75.0%;  Score 33;  DB 1;  Length 466;
  Best Local Similarity   87.5%;  Pred. No. 65;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0,  Gaps    0;

Qy       1 RLLAALCG 8
           || |||||
Db     175 RLPAALCG 182


RESULT  13
HYI2_AGRVI
ID   HYI2_AGRVI     STANDARD;      PRT;   467 AA.
AC   P25016;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Indoleacetamide hydrolase (EC 3.5.1.-) (IAH) (Indole-3-acetamide
DE   hydrolase).
GN   TA-IAAH.
OS   Agrobacterium vitis (Rhizobium vitis).
OG   Plasmid pTiTM4.
OC   Bacteria; Proteobacteria, alpha subdivision, Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=373;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TM4;
RX   MEDLINE=91329707; PubMed=1868204;
RA   Bonnard G., Vincent F., Otten L.;
RT   "Sequence of Agrobacterium tumefaciens biotype III auxin genes.";
RL   Plant Mol. Biol. 16:733-738(1991).
CC   -!- CATALYTIC ACTIVITY: INDOLE-3-ACETAMIDE (IAM) = INDOLE-3-
CC       ACETIC ACID (IAA).
CC   -!- PATHWAY: SECOND STEP IN THE BIOSYNTHESIS OF AUXINS FROM
CC       TRYPTOPHAN.
```

```
CC   -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
CC   -!- CAUTION: THE PLASMID PTITM4 CARRIES TWO T-REGIONS, THE TA AND
CC       TB REGION, BOTH OF WHICH HAVE AN IAAH GENE, WITH LOW HOMOLOGY
CC       BETWEEN THEM. ONLY THE TB-IAAH GENE SEEMS TO BE FUNCTIONAL.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56185; CAA39649.1; ALT_SEQ.
DR   PIR; S15452; HJAGI.
DR   PIR; S15001, S15001.
DR   InterPro; IPR000120, Amidase.
DR   Pfam; PF01425; Amidase; 1.
DR   PROSITE; PS00571; AMIDASES; 1.
KW   Hydrolase; Auxin biosynthesis; Crown gall tumor; T-DNA; Plasmid.
SQ   SEQUENCE   467 AA,  49941 MW,  8ABC04ADB58A1700 CRC64;


  Query Match             75.0%;  Score 33,  DB 1,  Length 467,
  Best Local Similarity   87.5%;  Pred. No. 66,
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0,  Gaps    0;

Qy       1 RLLAALCG 8
           || |||||
Db     175 RLPAALCG 182


RESULT  14
HYIN_AGRT4
ID   HYIN_AGRT4     STANDARD;      PRT;   467 AA.
AC   P03868;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Indoleacetamide hydrolase (EC 3.5.1.-) (IAH) (Indole-3-acetamide
DE   hydrolase).
GN   TMS2 OR IAAH.
OS   Agrobacterium tumefaciens (strain Ach5), and
OS   Agrobacterium tumefaciens.
OC   Plasmid pTiAch5, Plasmid pTiA6NC, and Plasmid pTi15955.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176298, 358;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=pTiA6NC;
RX   MEDLINE=84144041; PubMed=6366736;
RA   Sciaky D., Thomashow M.F.;
RT   "The sequence of the tms transcript 2 locus of the A. tumefaciens
RT   plasmid pTiA6 and characterization of the mutation in pTiA66 that is
RT   responsible for auxin attenuation";
RL   Nucleic Acids Res. 12:1447-1461(1984).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   PLASMID=pTi15955;
RA   Barker R.F., Idler K.B., Thompson D.V., Kemp J.D.;
RT   "Nucleotide sequence of the T-DNA region from the Agrobacterium
RT   tumefaciens octopine Ti plasmid pTi15955.";
RL   Plant Mol. Biol. 2:335-350(1983).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   PLASMID=pTiAch5;
RX   MEDLINE=84207942; PubMed=6327292;
RA   Gielen J., de Beuckeleer M., Seurinck J., Deboeck F., de Greve H.,
RA   Lemmers M., van Montagu M., Schell J.;
RT   "The complete nucleotide sequence of the TL-DNA of the Agrobacterium
RT   tumefaciens plasmid pTiAch5.";
RL   EMBO J. 3:835-846(1984).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   PLASMID=pTiA6NC;
RX   MEDLINE=84170374; PubMed=6584906;
RA   Klee H., Montoya A., Horodyski F., Lichtenstein C., Garfinkel D.,
RA   Fuller S., Flores C., Peschon J., Nester E., Gordon M.;
RT   "Nucleotide sequence of the tms genes of the pTiA6NC octopine Ti
RT   plasmid: two gene products involved in plant tumorigenesis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:1728-1732(1984).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   PLASMID=pTiAch5;
RA   Winans S.C., Zhu J., Oger P.M., Schrammeijer B., Hooykaas P.J.,
RA   Farrand S.K.;
RT   "Octopine-type Ti plasmid sequence.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: INDOLE-3-ACETAMIDE (IAM) + INDOLE-3-
CC       ACETIC ACID (IAA).
CC   -!- PATHWAY: SECOND STEP IN THE BIOSYNTHESIS OF AUXINS FROM
CC       TRYPTOPHAN.
CC   -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X00493; CAA25166.1; -.
DR   EMBL; X00409; CAA25116.1; -.
DR   EMBL; AF242881; AAF77122.1; -.
DR   EMBL; K02553; AAA92549.1; -.
DR   PIR; A04501; QQAGAT.
DR   PIR; B20966; B20966.
DR   PIR; S28586; S28586.
DR   InterPro; IPR000120, Amidase.
DR   Pfam; PF01425; Amidase; 1.
DR   PROSITE; PS00571; AMIDASES; 1.
KW   Hydrolase; Auxin biosynthesis; Crown gall tumor; T-DNA; Plasmid.
SQ   SEQUENCE   467 AA;  49805 MW;  2E0F1A3E72E4C1ED CRC64;


  Query Match             75.0%;  Score 33,  DB 1,  Length 467,
  Best Local Similarity   87.5%;  Pred. No. 66;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0,  Gaps    0;

Qy       1 RLLAALCG 8
           || |||||
Db     175 RLPAALCG 182


RESULT  15
RSEC_ECOLI
ID   RSEC_ECOLI     STANDARD;      PRT;   159 AA.
AC   P46187;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sigma-E factor regulatory protein rseC.
GN   RSEC OR B2570.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / W3110;
RA   Missiakas D., Raina S.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=K12;
RA   Connolly L., de Las Penas A., Gross C.A.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RA   Nashimoto H., Saito N.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- FUNCTION: SEEMS TO MODULATE THE ACTIVITY OF RPOE (SIGMA-E).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- SIMILARITY: TO H.INFLUENZAE H1589 AND H1850.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U37089; AAC45317.1; -.
DR   EMBL; U37455; AAC45320.1; -.
DR   EMBL; D64044; BAA10917.1; -.
DR   EMBL; AE000343; AAC75623.1; -.
DR   EcoGene; EG13178; rseC.
KW   Transmembrane; Inner membrane; Complete proteome.
FT   TRANSMEM     80    100       POTENTIAL.
FT   TRANSMEM    101    121       POTENTIAL.
SQ   SEQUENCE   159 AA;  16639 MW;  BCFD897ODF6277C0 CRC64;


  Query Match             72.7%;  Score 32;  DB 1;  Length 159;
  Best Local Similarity   85.7%;  Pred. No. 39;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 LAALCGA 9
           :||||||
Db     104 VAALCGA 110


Search completed. September  5, 2002, 15:31:44
Job time: 479 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:21:07 ; Search time 122.86 Seconds
(without alignments)
12.673 Million cell updates/sec

Title: US-09-744-804-41
Perfect score: 44
Sequence: 1 PLLAALCCA 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172934929 residues

Total number of hits satisfying chosen parameters. 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	100.0	335	4	Q9BTL9	Q9btl9 homo sapien
2	37	84.1	426	11	Q9WTS3	Q9wts3 mus musculu
3	37	84.1	464	11	Q9R1X9	Q9r1x9 mus musculu
4	36	81.8	144	16	Q9HX34	Q9hx34 pseudomonas
5	36	81.8	355	7	Q95601	Q95601 gallus gall
6	36	81.8	355	7	Q31400	Q31400 gallus gall
7	36	81.8	355	7	O46788	O46788 gallus gall
8	36	81.8	355	7	O46789	O46789 gallus gall
9	36	81.8	355	7	O46790	O46790 gallus gall
10	36	81.8	355	7	O46791	O46791 gallus gall
11	36	81.8	355	13	O73901	O73901 gallus gall
12	36	81.8	355	13	O73904	O73904 gallus gall
13	36	81.8	1513	4	Q9NXY3	Q9nxy3 homo sapien
14	35	79.5	230	16	Q9A7Y0	Q9a7y0 caulobacter
15	35	79.5	611	2	Q9L0K1	Q9l0k1 streptomyce
16	35	79.5	818	2	Q93H51	Q93h51 streptomyce
17	34	77.3	171	16	Q9A8J1	Q9a8j1 caulobacter
18	34	77.3	312	2	Q9L0Z3	Q9l0z3 streptomyce
19	34	77.3	376	16	Q9RWV5	Q9rwv5 deinococcus
20	34	77.3	466	5	Q9VQP2	Q9vqp2 drosophila
21	34	77.3	573	2	Q93MV6	Q93mv6 myxococcus
22	34	77.3	879	4	Q9NYX9	Q9nyx9 homo sapien
23	34	77.3	879	4	Q96BC3	Q96bc3 homo sapien
24	34	77.3	1699	4	Q13160	Q13160 homo sapien
25	34	77.3	1744	4	Q9UIP5	Q9uip5 homo sapien
26	34	77.3	1741	1	Q9NE85	Q9ne85 homo sapien
27	34	77.3	1744	4	Q13906	Q13906 homo sapien
28	33	75.0	28	11	Q9WTS1	Q9wts1 rattus norv
29	33	75.0	29	11	Q9JLR6	Q9jlr6 rattus norv
30	33	75.0	218	16	O07236	O07236 mycobacteri
31	33	75.0	235	10	Q9SN45	Q9sn45 arabidopsis
32	33	75.0	294	2	Q9R6Q6	Q9r6q6 agrobacteri
33	33	75.0	307	5	Q18327	Q18327 caenorhabdi
34	33	75.0	313	10	Q9SS71	Q9ss71 arabidopsis
35	33	75.0	334	2	Q93EV5	Q93ev5 streptomyce
36	33	75.0	347	2	Q9RL31	Q9rl31 streptomyce
37	33	75.0	461	2	Q9RBQ5	Q9rbq5 variovorax
38	33	75.0	466	2	Q9WWB5	Q9wwb5 agrobacteri
39	33	75.0	467	2	Q9R695	Q9r695 agrobacteri
40	33	75.0	467	2	Q9R712	Q9r712 agrobacteri
41	33	75.0	467	2	Q9WWE2	Q9wwe2 agrobacteri
42	33	75.0	467	2	P94205	P94205 agrobacteri
43	33	75.0	469	2	Q9X803	Q9x803 streptomyce
44	33	75.0	472	3	O59907	O59907 neurospora
45	33	75.0	473	2	Q93L17	Q93l17 burkholderi

ALIGNMENTS

RESULT 1
Q9BTL9
ID Q9BTL9 PRELIMINARY; PRT; 335 AA.
AC Q9BTL9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO MILK FAT GLOBULE-EGF FACTOR 8 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA.;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
DR EMBL; BC003610; AAH03610.1; -.
DR HSSP; P08709; 1BF9.
DR InterPro; IPR000561; EGF_like.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR000421; FA58_C.
DR PRINTS; PR00010; EGFBLOOD.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01285; FA58C_1; 1.
DR PROSITE; PS01286; FA58C_2; 2.
KW EGF like domain; Glycoprotein.
SQ SEQUENCE 335 AA; 37523 MW; 20E84E5DB78E9EB5 CRC64;

Query Match 100.0%; Score 44; DB 4; Length 335;
Best Local Similarity 100.0%; Pred No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 RLLAALCGA 9
           |||||||||
Db       5 RLLAALCGA 13


RESULT   2
Q9WTS3
ID   Q9WTS3       PRELIMINARY;      PRT:   426 AA.
AC   Q9WTS3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MILK FAT GLOBULE GLYCOPROTEIN MFG-E8 S (SIMILAR TO MILK FAT GLOBULE-
DE   EGF FACTOR 8 PROTEIN).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=MAMMARY GLAND;
RX   MEDLINE=99120894; PubMed=9920772;
RA   Oshima K., Aoki N., Negi M., Kishi M., Kitajima K., Matsuda T.;
RT   "Lactation-dependent expression of an mRNA splice variant with an exon
RT   for a multiply O-glycosylated domain of mouse milk fat globule
RT   glycoprotein MFG-E8.";
RL   Biochem. Biophys. Res. Commun. 254:522-528(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS
DR   EMBL; AB025280; BAA76486.1; -.
DR   EMBL; BC003904; AAH03904.1; -.
DR   EMBL; BC003892; AAH03892.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001438; EGF_11.
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR001092; HLH_dim.
DR   Ptam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00181; EGF; 2
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; 2
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   EGF-like domain; Glycoprotein.
SQ   SEQUENCE   426 AA;  47197 MW;  B182F4A2062B881A CRC64;


  Query Match             84.1%;  Score 37;  DB 11;  Length 426;
  Best Local Similarity   87.5%;  Pred. No. 40;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RLLAALCG 8
           |:|||||| 
Db       5 RVLAALCG 12


RESULT   3
Q9R1X9
ID   Q9R1X9       PRELIMINARY;      PRT:   463 AA.
AC   Q9R1X9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MILK FAT GLOBULE GLYCOPROTEIN MFG-E8 LONG FORM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=MAMMARY GLAND;
RX   MEDLINE=99120894; PubMed=9920772;
RA   Oshima K., Aoki N., Negi M., Kishi M., Kitajima K., Matsuda T.;
RT   "Lactation-dependent expression of an mRNA splice variant with an exon
RT   for a multiply O-glycosylated domain of mouse milk fat globule
RT   glycoprotein MFG-E8.";
RL   Biochem. Biophys. Res. Commun. 254:522-528(1999).
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR   EMBL; AB021130; BAA35180.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001438; EGF_11.
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR001092; HLH_dim.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; 2
DR   PROSITE; PS01285; FA58C_1; 2
DR   PROSITE; PS01286; FA58C_2; 2.
KW   EGF-like domain; Glycoprotein.
SQ   SEQUENCE   463 AA;  51269 MW;  D719D2BE090F6427 CRC64;


  Query Match             84.1%;  Score 37;  DB 11;  Length 463;
  Best Local Similarity   87.5%;  Pred. No. 44;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RLLAALCG 8
           |:|||||| 
Db       5 RVLAALCG 12


RESULT   4
Q9HX34
ID   Q9HX34       PRELIMINARY;      PRT:   143 AA.
AC   Q9HX34;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   HYPOTHETICAL PROTEIN PA3986.
GN   PA3986.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004816; AAG07373.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   143 AA;  15937 MW;  8845B066D3813B03 CRC64;


  Query Match             81.8%;  Score 36;  DB 16;  Length 143;
  Best Local Similarity   77.8%;  Pred. No. 24;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy      1 RLLAALCGA 9
          :|| |||||
Db    108 QIICALCGA 116


RESULT   5
Q95601
ID   Q95601      PRELIMINARY;      PRT;   355 AA.
AC   Q95601;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MAJOR HISTOCOMPATIBILITY COMPLEX CLASS I GLYCOPROTEIN HAPLOTYPE B21.
GN   BFIV21 OR B-FIV.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98347411; PubMed=7621880;
RA   Fulton J.E., Thacker E.L., Bacon L.D., Hunt H.D.;
RT   "Functional analysis of avian class I (BFIV) glycoproteins by epitope
RT   tagging and mutagenesis in vitro.";
RL   Eur. J. Immunol. 25:2069-2076(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=15I5 B21 CONGENIC;
RA   Hunt H.D.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC       IMMUNE SYSTEM (BY SIMILARITY).
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN) (BY SIMILARITY).
DR   EMBL; S78682; AAB34945.1; -.
DR   EMBL; AF013493; AAC17590.1; -.
DR   HSSP; P13599; 3FRU.
DR   InterPro; IPR003597; Ig_c1.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001039; MHC_I.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00129; MHC_I; 1.
DR   ProDom; PD000050; MHC_I; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; UNKNOWN_1.
KW   Glycoprotein; MHC; Transmembrane.
SQ   SEQUENCE   355 AA;  38960 MW;  A751A09C46D94EE9 CRC64;


  Query Match             81.8%;  Score 36;  DB 7;  Length 355;
  Best Local Similarity   87.5%;  Pred. No. 52;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0,  Gaps    0,

Qy      2 LLAALCGA 9
          ||||:|||
Db     12 LLAAVCGA 19


RESULT   6
Q31400
ID   Q31400      PRELIMINARY;      PRT;   355 AA.
AC   Q31400;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MAJOR CLASS I GLYCOPROTEIN PRECURSOR.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB; TISSUE=BURSA;
RA   Walny H., Avila D., Riegert P., Salomonsen J., Vilbois F., Wiles M.,
RA   Kaufman J.;
RT   "Peptide motifs for the dominantly expressed class I molecule of the
RT   chicken Major Histocompatibility Complex and resistance to infectious
RT   pathogens.";
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC       IMMUNE SYSTEM (BY SIMILARITY).
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN) (BY SIMILARITY).
DR   EMBL; Z54326; CAA91118.1; -.
DR   HSSP; P13599; 3FRU.
DR   InterPro; IPR003597; Ig_c1.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001039; MHC_I.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00129; MHC_I; 1.
DR   ProDom; PD000050; MHC_I; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; UNKNOWN_1.
KW   Glycoprotein; Signal; Transmembrane.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    355       MAJOR CLASS I GLYCOPROTEIN.
SQ   SEQUENCE   [illegible] AA;  [illegible] MW;  [illegible] CRC64;


  Query Match             81.8%;  Score 36;  DB 7;  Length 355;
  Best Local Similarity   87.5%;  Pred. No. 52;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0,  Gaps    0;

Qy      2 LLAALCGA 9
          ||||:|||
Db     12 LLAAVCGA 19


RESULT   7
O46788
ID   O46788      PRELIMINARY;      PRT;   355 AA.
AC   O46788;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MHC CLASS I GLYCOPROTEIN.
GN   B-FIV.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN 15I5 B5 CONGENIC;
RX   MEDLINE 98221119; PubMed 9553152;
RA   Hunt H.D., Fulton J.E.;
RT   "Analysis of polymorphisms in the major expressed class I locus (B-
RT   FIV) of the chicken.";
RL   Immunogenetics 47:456-467(1998).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC       IMMUNE SYSTEM (BY SIMILARITY).
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA 2-
CC       MICROGLOBULIN) (BY SIMILARITY).
DR   EMBL; AF013491; AAC17588.1; -.
DR   HSSP; P13599; 3FRU.
DR   InterPro; IPR003597; Ig_c1.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001039; MHC_I.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00129; MHC_I; 1.
DR   ProDom; PD000050; MHC_I; 1.
DR   SMART; SM00407; IGc1; 1.
```

```
DR   PROSITE; PS00290; IG_MHC; UNKNOWN_1.
KW   Glycoprotein; Transmembrane.
SQ   SEQUENCE   355 AA;  38953 MW;  F7BCF9B88F61582E CRC64;


  Query Match             81.8%;  Score 36;  DB 7;  Length 355;
  Best Local Similarity   87.5%;  Pred. No. 52;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
           ||||:|||
Db      12 LLAAVCGA 19


RESULT   8
O46789
ID   O46789      PRELIMINARY;      PRT;   355 AA
AC   O46789;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MHC CLASS I GLYCOPROTEIN.
GN   B-FIV.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=15I5 B2 CONGENIC;
RX   MEDLINE=98221119; PubMed=9553152;
RA   Hunt H.D., Fulton J.E.;
RT   "Analysis of polymorphisms in the major expressed class I locus (B-
RT   FIV) of the chicken.";
RL   Immunogenetics 47:456-467(1998).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC       IMMUNE SYSTEM (BY SIMILARITY).
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN) (BY SIMILARITY).
DR   EMBL; AF013492; AAC17589.1; -.
DR   InterPro; IPR003597; Ig_cl.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001039; MHC_I.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00129; MHC_I; 1.
DR   ProDom; PD000050; MHC_I; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; UNKNOWN_1.
KW   Glycoprotein; Transmembrane.
SQ   SEQUENCE   355 AA;  39049 MW;  3DEC7E5D713EDEE0 CRC64;


  Query Match             81.8%;  Score 36;  DB 7;  Length 355;
  Best Local Similarity   87.5%;  Pred. No. 52;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
           ||||:|||
Db      12 LLAAVCGA 19


RESULT   9
O46790
ID   O46790      PRELIMINARY;      PRT;   355 AA.
AC   O46790;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MHC CLASS I GLYCOPROTEIN.
GN   B-FIV.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=I515 B13 CONGENIC;
RX   MEDLINE=98221119; PubMed=9553152;
RA   Hunt H.D., Fulton J.E.;
RT   "Analysis of polymorphisms in the major expressed class I locus (B-
RT   FIV) of the chicken.";
RL   Immunogenetics 47:456-467(1998).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC       IMMUNE SYSTEM (BY SIMILARITY).
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN) (BY SIMILARITY).
DR   EMBL; AF013494; AAC17591.1; -.
DR   InterPro; IPR003597; Ig_cl.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001039; MHC_I.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00129; MHC_I; 1.
DR   ProDom; PD000050; MHC_I; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; UNKNOWN_1.
KW   Glycoprotein; Transmembrane.
SQ   SEQUENCE   355 AA;  39398 MW;  8B08F669AFF5AF37 CRC64;


  Query Match             81.8%;  Score 36;  DB 7;  Length 355;
  Best Local Similarity   87.5%;  Pred. No. 52;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
           ||||:|||
Db      12 LLAAVCGA 19


RESULT  10
O46791
ID   O46791      PRELIMINARY;      PRT;   355 AA.
AC   O46791;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MHC CLASS I GLYCOPROTEIN.
GN   B-FIV.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B17 CONGENIC;
RX   MEDLINE=98221119; PubMed=9553152;
RA   Hunt H.D., Fulton J.E.;
RT   "Analysis of polymorphisms in the major expressed class I locus (B-
RT   FIV) of the chicken.";
RL   Immunogenetics 47:456-467(1998).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC       IMMUNE SYSTEM (BY SIMILARITY).
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN) (BY SIMILARITY).
DR   EMBL; AF013495; AAC17592.1; -.
DR   HSSP; P13599; 3FRU.
DR   InterPro; IPR003597; Ig_cl.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001039; MHC_I.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00129; MHC_I; 1.
DR   ProDom; PD000050; MHC_I; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; UNKNOWN_1.
```

```
KW   Glycoprotein; Transmembrane.
SQ   SEQUENCE   355 AA;  39170 MW;  5638297709A6AAFB CRC64;


  Query Match             81.8%;  Score 36;  DB 7;  Length 355;
  Best Local Similarity   87.5%;  Pred. No. 52;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
            ||||:|||
Db      12 LLAAVCGA 19


RESULT  11
O73901
ID   O73901      PRELIMINARY;      PRT;   355 AA.
AC   O73901;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   B LOCUS F ALPHA CHAIN 2.
GN   BFA2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89030642; PubMed=3141149;
RA   Guillemot F., Billaut A., Pourquie O., Behar G., Chausse A.M.,
RA   Zoorob R., Kreibich G., Auffray C.;
RT   "A molecular map of the chicken major histocompatibility complex: the
RT   class II beta genes are closely linked to the class I genes and the
RT   nuclear organizer.";
RL   EMBO J. 7:2775-2785(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Milne S., Kaufman J., Beck S.;
RT   "DNA sequencing and analysis of the chicken major histocompatibility
RT   complex.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC       IMMUNE SYSTEM (BY SIMILARITY).
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN) (BY SIMILARITY).
DR   EMBL; AL023516; CAA18969.1; -.
DR   HSSP; Q30201; 1A6Z.
DR   InterPro; IPR003597; Ig_c1.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001039; MHC_I.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00129; MHC_I; 1.
DR   ProDom; PD000050; MHC_I; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; UNKNOWN_1.
KW   Glycoprotein; Transmembrane.
SQ   SEQUENCE   355 AA;  38695 MW;  648D24F592FD98AB CRC64;


  Query Match             81.8%;  Score 36;  DB 13;  Length 355;
  Best Local Similarity   87.5%;  Pred. No. 52;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
            ||||:|||
Db      12 LLAAVCGA 19


RESULT  12
O73904
ID   O73904      PRELIMINARY;      PRT;   355 AA
AC   O73904;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   B LOCUS F ALPHA CHAIN 1.
GN   BFA1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89030642; PubMed=3141149;
RA   Guillemot F., Billaut A., Pourquie O., Behar G., Chausse A.M.,
RA   Zoorob R., Kreibich G., Auffray C.;
RT   "A molecular map of the chicken major histocompatibility complex: the
RT   class II beta genes are closely linked to the class I genes and the
RT   nuclear organizer.";
RL   EMBO J. 7:2775-2785(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Milne S., Kaufman J., Beck S.;
RT   "DNA sequencing and analysis of the chicken major histocompatibility
RT   complex.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC       IMMUNE SYSTEM (BY SIMILARITY).
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN) (BY SIMILARITY).
DR   EMBL; AL023516; CAA18972.1; -.
DR   HSSP; P13599; 3FRU.
DR   InterPro; IPR003597; Ig_c1.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001039; MHC_I.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00129; MHC_I; 1.
DR   ProDom; PD000050; MHC_I; 1.
DR   SMART; SM00407; IGc1; 1.
DR   PROSITE; PS00290; IG_MHC; UNKNOWN_1.
KW   Glycoprotein; Transmembrane.
SQ   SEQUENCE   355 AA;  39200 MW;  D8D987452F18467A CRC64;


  Query Match             81.8%;  Score 36;  DB 13;  Length 355;
  Best Local Similarity   87.5%;  Pred. No. 52;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
            ||||:|||
Db      12 LLAAVCGA 19


RESULT  13
Q9NXY3
ID   Q9NXY3      PRELIMINARY;      PRT;  3513 AA.
AC   Q9NXY3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FLJ00002 PROTEIN (FRAGMENT).
GN   FLJ00002.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SPLEEN;
RA   Ohara O., Kikuno R., Nagase T., Okumura K.;
RT   "The nucleotide sequence of a long cDNA clone isolated from human
RT   spleen.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK000002; BAA92227.1; -.
```

```
DR   BSSP; P13569; 1NBD.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR001140; ABC_transporter_tmem.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   Pfam; PF00664; ABC_membrane; 2.
DR   Pfam; PF00005; ABC_tran; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_?
KW   ATP-binding.
FT   NON_TER       1       1
SQ   SEQUENCE   1513 AA;  163831 MW;  FBAA7AABDE68AE3C CRC64;


  Query Match             81.8%;  Score 36;  DB 4;  Length 1513;
  Best Local Similarity   77.8%;  Pred. No. 2e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 RLLAALCGA 9
           |||| |||:
Db      24 RLLAQLCGS 32


RESULT  14
Q9A7Y0
ID   Q9A7Y0      PRELIMINARY;      PRT;   230 AA.
AC   Q9A7Y0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL PROTEIN CC1589.
GN   CC1589.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=69394;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005833; AAK23568.1; -.
DR   TIGR; CC1589; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   230 AA;  24454 MW;  DE84CD3F28E54BD1 CRC64;


  Query Match             79.5%;  Score 35;  DB 16;  Length 230;
  Best Local Similarity   87.5%;  Pred. No. 54;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
           |||| |||
Db       6 LLAAFCGA 13


RESULT  15
Q9L0K1
ID   Q9L0K1      PRELIMINARY;      PRT;   611 AA.
AC   Q9L0K1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE MEMBRANE PROTEIN.
GN   SCD40A.10C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RT   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RT   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL161691; CAB81855.1; -.
DR   InterPro; IPR003838; DUF214.
DR   Pfam; PF02687; DUF214; 1.
SQ   SEQUENCE   611 AA;  61744 MW;  8F959720773AEDF9 CRC64;


  Query Match             79.5%;  Score 35;  DB 2;  Length 611;
  Best Local Similarity   87.5%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LLAALCGA 9
           || |||||
Db     140 LLVALCGA 147


Search completed: September  5, 2002, 15:31:08
Job time: 473 sec
```